AD-A286 937

DMDC Report No. 97-028
December 1997

DMDC Sample Planning Tool: User's Manual (Version 1.2)

98-00005

Defense Manpower Data Center
Survey & Program Evaluation Division
1600 Wilson Boulevard, Suite 400
Arlington, VA 22209-2593

Best
Available
Copy

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters, Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (*Leave Blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	03 Dec 97	Final (29 Sep 94 -- 31Oct 97)

4. TITLE AND SUBTITLE

DMDC Sample Planning Tool User's Manual (Version 1.2)

5. FUNDING NUMBERS

C - DASW01-94-H-0002
(DO No. 0001)

6. AUTHOR(S)

Jill D. Kavee & Robert E. Mason

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Research Triangle Institute, 3040 Cornwallis Road, PO BOX 12194, Research Triangle Park, NC 27709-2194

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Defense Manpower Data Center (DMDC), Survey & Program Evaluation Division, 1600 Wilson Boulevard Suite 400, Arlington, VA 22209-2593

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

97-028

11. SUPPLEMENTARY NOTES

Technical Monitors: Richard A. Riemer & Timothy W. Elig
Report is accompanied by the Tool which is written in Visual Basic and executed in Microsoft Access 2.0

12a. DISTRIBUTION/AVAILABILITYSTATEMENT

Approved for public release;
distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

This report describes the Sample Planning Tool (Tool) developed for surveys conducted by the Defense Manpower Data Center (DMDC). The approach implemented by the Tool is particularly suited for situations where there is a great depth of information (such as in personnel databases) available on the population to be sampled. The software is designed to produce optimal sample designs for the general class of designs with cost models that are restricted to mail data collection procedures and variance models that are restricted to stratified random sampling design. Features of the Tool assist in: constructing and stratifying the sampling frame, constructing cost and variance models, defining reporting domains to provide the basis for specifying the precision requirements for the surveys, and specifying and imposing precision requirements. With this information the Tool computes the minimum cost allocation of the sample that will satisfy the precision requirements. The mathematical basis for the Tool is provided by the Karush-Kuhn-Tucker necessary conditions for function minimization (Kuhn & Tucker, 1951) as described by Chromy (1987). The intended users of the Tool are sampling statisticians and other analysts familiar with sampling theory. The Tool is written in Visual Basic and executed in Microsoft Access 2.0.

14. SUBJECT TERMS	15. NUMBER OF PAGES
surveys, survey methods, statistics, survey sampling	90
	16. PRICE CODE

17. SECURITY CLASSIFICATON OF REPORT	18. SECURITY CLASSIFICATION OF THE PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

SF298 (Rev 7/95)

For additional copies of the application software and this report,

contact:

Defense Technical Information Center
ATTN: DTIC-BRR
Defense Document Information Center
8725 John J. Kingman Rd., Suite #0944
Ft. Belvoir, VA 22060-6218

(703) 767-8274

Ask for Report by AD # M000772

DMDC Report No. 97-028
December 1997

D..DC SAMPLE PLANNING TOOL

USER'S MANUAL (VERSION 1.2)

Jill D. and Robert E. Mason
Rese.. a Triangle Institute

Technical Monitors:
Richard A. Riemer and Timothy W. Elig

Defense Manpower Data Center
Survey & Program Evaluation Division
1600 Wilson Boulevard, Suite 400, Arlington, VA 22209-2593

Executive Summary

This report describes the Sample Planning Tool (Tool) developed by Research Triangle Institute (RTI) for sample designs for surveys conducted by the Defense Manpower Data Center (DMDC). The software is designed to produce optimal sample designs for stratified random samples. The approach implemented by the Tool is particularly suited to the depth of information available to DMDC when survey populations are defined.

DMDC has a wealth of information about military personnel and their families to assist in the design of survey samples. Information is available to segment the population into specific groups that are expected to have differing response rates to surveys and prevalence rates associated with survey questions. Segmenting the population and sample based on this information allows for more precise survey estimates than would be afforded by a simple random sample. Furthermore, setting precision requirements for sets of these segments assures DMDC that the survey results will provide useful information for addressing policy questions. The Tool incorporates information about expected response rates, prevalence rates, precision requirements, and the population distribution to produce an optimal sample.

The Tool was developed as part of the design effort for the 1995 Sexual Harassment Survey and the 1996 Equal Opportunity Survey. DMDC has also used the Tool to design the samples for a number of other surveys: the 1996 Survey of Retired Military Personnel, the 1997 Junior Enlisted Spouse Survey, the DoD Financial Services Survey, the Survey of Parents' Opinions on Department of Defense Domestic Dependent Elementary and Secondary Schools, the Survey of Parents' Opinions on Local Schools.

The Tool is written in Visual Basic and executed in Microsoft Access 2.0. The Tool does have some sampling theory limitations. For example, the Tool is not suited for cluster designs (e.g., a sample of personnel from a sample of bases in which case the sample bases are considered clusters).

A formal mathematical procedure based on Karush-Kuhn-Tucker theory is used to determine the sample size and allocation. The procedure involved developing equations to describe the variance of the sample estimates and the variable survey costs, then simultaneously solving the equations subject to the (inequality) precision requirements. The obtained solution is unique and is the sample allocation that jointly satisfied the precision requirements for the least cost. The allocation procedure was first described by Chromy (1987).

This report assumes a background in sampling theory. The intended users of the Tool are sampling statisticians and other analysts familiar with sampling theory. In addition, users of the Tool need to have a firm understanding of the demographics of the target population and the analytic goals of the survey effort.

Table of Contents

Table of Contents (continued)

Appendices

Introduction

This document describes software developed by the Research Triangle Institute to assist the Defense Manpower Data Center (DMDC) in the development of sampling designs. This software is referred to as the Sample Planning Tool. The Sample Planning Tool was developed under the **1994/1995 Status of the Armed Forces Surveys (SAFS)** contract specifically to assist in the design of **the 1995 Sexual Harrassment Survey (SHS)** and the **1996 Equal Opportunity Survey (EOS).** Although developed specifically for these surveys, this version of the software is applicable to the general class of designs with cost models that are restricted to mail data collection procedures and variance models that are restricted to stratified random sampling design. Many DMDC surveys are of this general type of design. Features of the **Sample Planning Tool** assist in:

- constructing and stratifying the sampling frame,
- constructing cost and variance models,
- defining the reporting domains to provide the basis for specifying the precision requirements for the surveys, and
- specifying and imposing the precision requirements.

In this version of the Tool, cost models are restricted to mail data collection procedures and variance models are restricted to stratified random sampling designs.

With the above information the Tool computes the minimum cost allocation of the sample that will satisfy the precision requirements. The mathematical basis for the Tool is provided by the Karush-Kuhn-Tucker necessary conditions for function minimization (Kuhn & Tucker, 1951) as described by Chromy (1987). Appendix C of this manual contains a paper which gives details of the algorithm specific to the surveys in SAFS.

The Tool is designed to assist with the development of survey designs. Other activities are also a part of any probability-based survey, such as sample weight construction, nonresponse weight adjustments, and estimation. The user is referred to the methodology reports for the 1995 SHS (Mason, Kavee, Wheeless, George, Riemer, & Elig, 1996) and the 1996 EOS (Wheeless, Mason, Kavee, Riemer, & Elig, 1997) for examples specific to surveys of U.S. military personnel. A general discussion of survey design, weighting, and estimation can be found in sampling texts such as Cochran (1977) and Wolter (1985).

The Tool also generates summary reports that describe the salient features of the design and the sample allocation. The reports serve to assist both in the development of the sampling design and in its final documentation.

The Tool is written in **Visual Basic** and executed in **Microsoft Access 2.0**. This manual has been written for persons with a modest working knowledge of Access. In the vernacular of Access, program **objects** may be tables, queries, forms, reports, macros, or modules. The Tool user interface consists of an ordered set of forms, which can be thought of as screens through which the program communicates with the user and vice versa. However, users familiar with Access can access any of the objects that comprise the Tool.

Further, Tool objects can be exported to and imported from other Microsoft products that support object linking and embedding. For example, Access table objects can be exported to Excel for additional analysis and Access report objects can be exported for inclusion in reports being written in Word. As a cautionary note, the current version of the Tool makes no checks of table layouts and field definitions. The presumption is that the tables have been generated by the Tool. Importing tables from other sources, such as Excel, for use in the Tool may result in run-time errors if the table layout and format are not exactly that which is expected in the Visual Basic code.

This version of the Tool does not modify the standard Microsoft Access 2.0 menus or toolbars. These remain as they are configured by Access with Access-defined functions. The user is guided through the Tool by forms that have the word *Menu* in the window bar of the form rather than by the menus.

Each form in the Tool is discussed on at least one page in this document. The actual screen presentations are included with the text to assist with the explanation. Some general rules are applied to aid in reading the manual as follows.

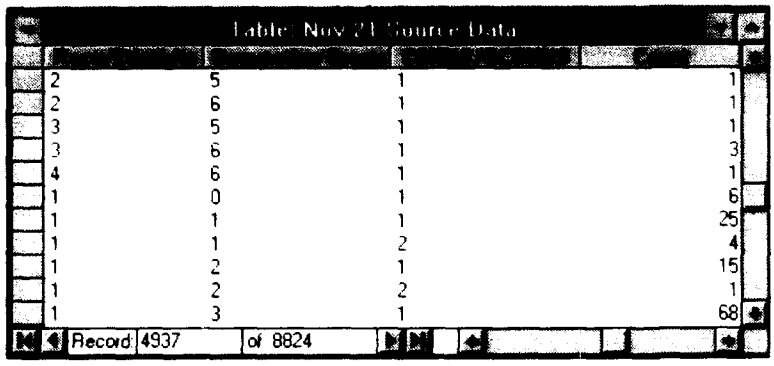
- Table and data set names are identified by quotation marks, for example, "Cost Data". The contents of the significant tables discussed in the report are provided in Appendix A.
- Form names are written in italicized, sentence-case form, for example *Define Stratum Levels*.
- Objects within a form, such as list boxes and controls, are written in italicized, upper-case form, for example, *CONTINUE*.

Initializing The Tool

To initialize the Tool two types of information must be supplied. The information must be created and stored in table-form as a part of the Tool database before running the Tool itself.

The first type of information is the "Source Data" which can be supplied in one or more tables. The tables, in the aggregate, contain all of the relevant variables and variable values that are needed to construct the strata and to define the reporting domains. The character string "Source Data" must appear at the end of the otherwise arbitrary names given to the tables. The variables appear as fields (or columns) in the table with the variable values being the entries in each field. The field names (or column headings) assigned by the user become the variable names used by the Tool in all subsequent steps.

The last field in the table is labeled *COUNT*. This field contains the numbers of population units to which the variable values listed in the row of the table apply. The "Source Data" table must contain the *COUNT* field and at least one variable field. This version of the tool is programmed to handle only categorical variables which can be coded using either numeric or character strings. The variable value or code fields are text fields of sufficient length to accommodate the codes. The *COUNT* field is numeric, using long integers (4 byte words).



2	5	1	1
2	6	1	1
3	5	1	1
3	6	1	3
4	6	1	1
1	0	1	6
1	1	1	25
1	1	2	4
1	2	1	15
1	2	2	1
1	3	1	68

Record: 4937 of 8824

In this example of a section of a "Source Data" table, three variables are shown in addition to the counts of individuals. The tool from this point on knows the variables by the field names (column headings) used in the table.

The second type of information consists of tables that identify the variable codes listed in the "Source Data" table or tables. The codes for each variable appear in a separate table. The code tables are named "List Of (variable name) Codes", where the variable name is identically the field name used in the "Source Data" table. Each table is exhaustive in the sense that every allowable value or code for the variable in question appears in the table. Conversely, the Tool assumes that any variable value that appears in the "Source Data" table and not listed in the appropriate code table is an invalid code. Codes not identified in the appropriate code table are considered unknown for purposes of constructing strata and defining reporting domains.

Each code table consists of two fields. The first field in the table is labeled with the same name as the name assigned to the field that contains the corresponding variable in the "Source Data" table. The field contains a description of the variable value or code that appears in the same row of the second field in the table. These code labels will appear in most reports generated by the Tool. The label of the second field in the table repeats the title of the first field with the word "Code" added.

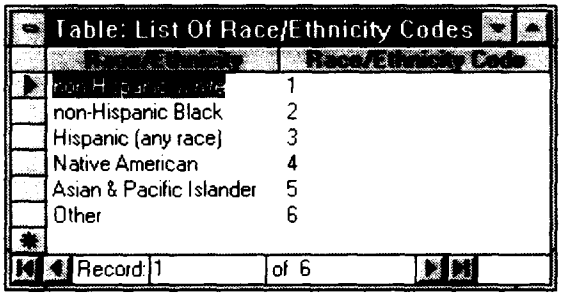


Table: List Of Race/Ethnicity Codes

[illegible]	Race/Ethnicity Code
[illegible]	1
non-Hispanic Black	2
Hispanic (any race)	3
Native American	4
Asian & Pacific Islander	5
Other	6

Record: 1 of 6

The example code table provides the exhaustive list of variable values or codes assigned to the Race/Ethnicity variable in the above "Source Data" table segment. Note that the field name Race/Ethnicity assigned to the variable in the "Source Data" table is repeated in the code table name and both fields in the code tab'e.

Several code tables have been developed in the course of using the tool for the SAFS and are already included in the Tool database. The list of available code tables can be viewed by loading the tool and unhiding the Database window. The Database window is unhidden by selecting the menu item Unhide from the Access *WINDOW* menu, if the *WINDOW* menu is showing, or from the *FILE* menu if the *WINDOW* menu is not showing.

Loading The Tool

The diskette accompanying this manual contains the self-extracting file MSTRTOOL.EXE. Create a hard disk directory to contain the Tool and make this directory the active directory. Insert the diskette in a floppy drive, say drive A, and execute the DOS command A:\MSTRTOOL from the active directory.

The file MSTRTOOL.MDB will be created in the active directory. To use the Tool, first copy MSTRTOOL.MDB to a working directory giving it a new name in the process, for example, FORMD.MDB. Keeping the extension .MDB on all copies of the Tool allows Access 2.0 to be opened and the correct copy of the Tool to be loaded simply by double clicking the appropriate file name in Windows *FILE MANAGER*. Alternately Access 2.0 can be opened and the appropriate copy of the Tool loaded from within Access.

Constructing A Source Data Table

Prepare the source data as a worksheet in a spreadsheet format recognized by Access (such as Excel) or as a fixed field or delimited ASCII file. Open Access 2.0 and load an appropriately re-named copy of MSTRTOOL.MDB. The tool opens showing an identification screen which is replaced in a few seconds by the form *Sampling Tool Menu*. Close the *Sampling Tool Menu* using the *CONTROL-MENU BOX* located on the extreme upper left corner of the form window (a standard Windows control; see The Parts of a Window, Microsoft Windows User's Guide, Microsoft Corporation, page 8).

Click the Access *FILE* menu item *UNHIDE* and unhide the *DATABASE* window. This action changes the Access *FILE* menu items, which now include the item *IMPORT...*. Click *IMPORT...* and follow the instructions in the Access dialog boxes to load the source data. If more than one file is needed for all of the source data, they can each be imported at this point.

When the source data has been added to the Tool database each file will be identified by its capitalized DOS file name under *TABLES* in the *DATABASE* window. Highlight the imported source data tables in turn, click on the Access *FILE* menu item *RENAME* and rename the table or tables following the "Source Data" Tool name conventions described previously.

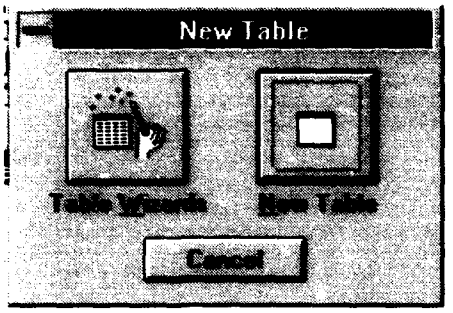
The field names in the tables can be added or edited and the tables formatted by clicking on the table names and opening the tables in *DESIGN VIEW* from the Tool *DATABASE* window. Fields containing the variable values or codes are formatted as text fields with field sizes large enough to accommodate the variable value code with the most characters. The *COUNT* field is formatted as a long integer numeric field.

Return control to the Tool by clicking *FORMS* on the *DATABASE* window and double clicking the form *Sampling Tool Menu*.

Constructing A Code Table

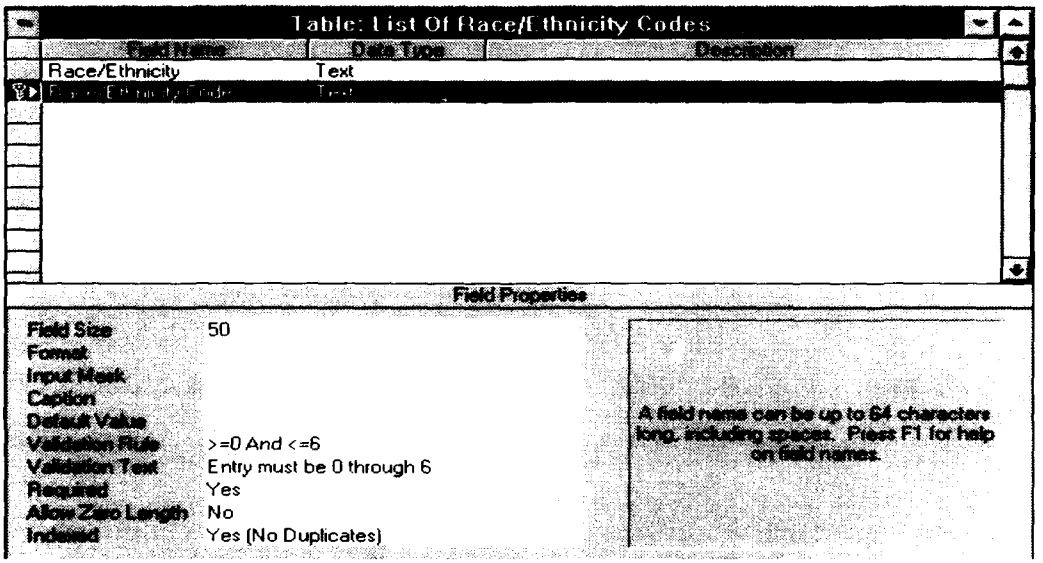
Code tables can be imported following the same procedure described above for the "Source Data" tables. However, the code tables are small enough that it is likely easier to construct these using Access directly.

Unhide the *DATABASE* window, if it is hidden, using the Access *WINDOWS* menu or *FILE* menu as appropriate. Click *TABLES* and then click *NEW* on the *DATABASE* window.



Access responds with the choice between using a wizard or creating a table without assistance. Click *NEW TABLE*. Access will open a new table in design view. Under Field Name, first key the variable name assigned in the "Source Data" table, and then key the same name followed by the word Code. Assign both fields a data type of Text. Set other field characteristics as desired. Save the table using the *SAVE AS* item on the Access *FILE* menu and the naming convention discussed earlier.

The example below shows the "List Of Race/Ethnicity Codes" table open in design view with the Race/Ethnicity Code field highlighted. Note that this field has been designated as key, indicated by the outline of a key to the left of the field name. Designating a field as key, among other things prevents the keying of any duplicate entries in the field. This feature ensures that the same code has not been assigned more than once. Further, the table will always be in sort by the values contained in the key field. Other useful (but not essential) field properties are listed in the bottom part of the figure.



Entries in the two fields are simply keyed into the table after switching the table to datasheet design view.

New Features In Version 1.2

Between 1995 and 1997, DMDC has used version 1.1 of the Tool to design several surveys. This experience has suggested several enhancements that have been incorporated in version 1.2. DMDC's work in this period also identified some minor problems in version 1.1 that have been corrected in version 1.2. The enhancements listed below have not been incorporated into the subsequent sections of the Sampling Tool User's Manual.

Enhancements

- In some DMDC applications, the time interval between the development and approval of the sampling design and selection of the sample is such that the source information used to construct the sampling frame and reporting domains will have been updated in the interim. Version 1.1 of the Tool necessitated repeating the steps associated with constructing the strata and defining the domains in order to incorporate the updated source information. In version 1.2, new command buttons have been added to the forms *Construct Strata* and *Construct Stratum/Domain Counts* that will incorporate the updated information without the necessity of repeating the previous work. After the relevant form is invoked, press the USE PREVIOUS STRATUM DEFINITIONS command button to process the new frame. Pressing DONE will close the form and save the new counts to the corresponding tables.

- The precision requirements for a survey are specified in terms of the confidence interval half-widths to be associated with sample estimates of the proportions of individuals in specified domains that exhibit arbitrary sets of characteristics. The domain proportions are referred to as prevalence estimates in the Tool documentation. In general, some of the domains defined for this purpose may be combinations of others. Under this circumstance a consistent value of the prevalence estimate for the combined domain is computed as the weighted average of the component domain prevalence estimates, where the weights are the relevant domain sizes. Version 1.1 of the Tool provides the user with no assistance in computing consistent values of the prevalence estimates. A new form, *Combine Domains*, has been added to version 1.2 that allows the user to combine previously defined domains and that will compute consistent values of the proportions for the resulting combinations. This form is invoked from the *Reporting Domains Menu*. After opening the form, select the domains to be combined to create the new domain using the SELECT command button. Press SELECTION COMPLETE once all of the domains have been selected. The DONE button calculates the new domain sizes and prevalence estimates, and closes the form.

- In a typical application, DMDC data processing personnel are responsible for selecting samples of specified sizes from within each of the design strata. Version 1.2 of the Tool generates a "lookup" table for communicating the relevant design specifications to the DP programmers. The new form *Export Lookup Table* is invoked from the *Sample Allocation Menu*. After pressing the CONTRUCT TABLE command button, this form generates a table containing the stratum identifiers, stratum sizes, stratum-level sample sizes, and the variable names and variable values (i.e., codes) that define the strata. The lookup table can be exported to a file using the Access Export utility after the DONE command button has been pressed.

- In constructing strata, the user specifies the variables to be used to define the dimensions of stratification and the variable values to be used to define the levels of stratification within each dimension. Not infrequently some of the initially specified strata turn out to be too

small to support a reasonable sample allocation and are combined with others. The user provides instructions for collapsing strata using the interface provided in the *Collapse Strata* form. This form has been enhanced in version 1.2 of the Tool to display the descriptive label or labels of the combined strata generated as a result of each combine instruction once the COLLAPSE STRATA button has been pressed. After viewing the label, the user can accept or reject the result. Rejecting the result automatically undoes the associated combine instruction allowing it to be re-formulated.

- Response rate information is supplied to the Tool is two pieces through the *Response Rate form*. The first piece is the expected response rate for each of the design strata. The second piece is the proportion of the stratum-level response rate that is expected to be obtained on each mailing. In a typical application, the stratum-level response rates are likely to be different in different strata while the proportions obtained at each mailing are likely to be the same in all strata. Version 1.2 of the Tool displays the previously keyed proportions so that they can be entered again without the necessity of re-keying them.
- The Tool *Menus* list the various steps in the design process. Version 1.2 contains some new menu items and re-orders others to provide a more logical sequencing. Also, some of the version 1.2 forms contain new command buttons and the labeling of some of the controls on some of the forms has been changed to better describe the intended functionality.

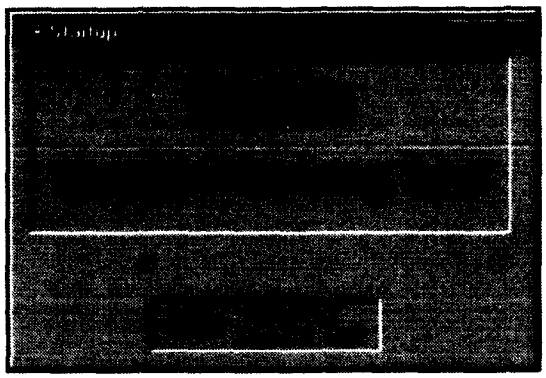
Corrections

- The calculation of stratum per unit cost coefficients has been corrected.
- The inconsistency in reporting the size of the "Unknown" stratum has been removed.
- Field widths for population and sample size totals provided on various reports have been expanded to accommodate eight digits.

Other Features

- The Tool was developed using Windows 3.x and the illustrations provided in this manual were produced using that operating system. The illustrations in the manual will have some minor differences with what appears on the screen when viewed with a more recent Windows release.
- The Tool was developed using Access 2.0 which is not compatible with more recent Access releases.

Using The Tool

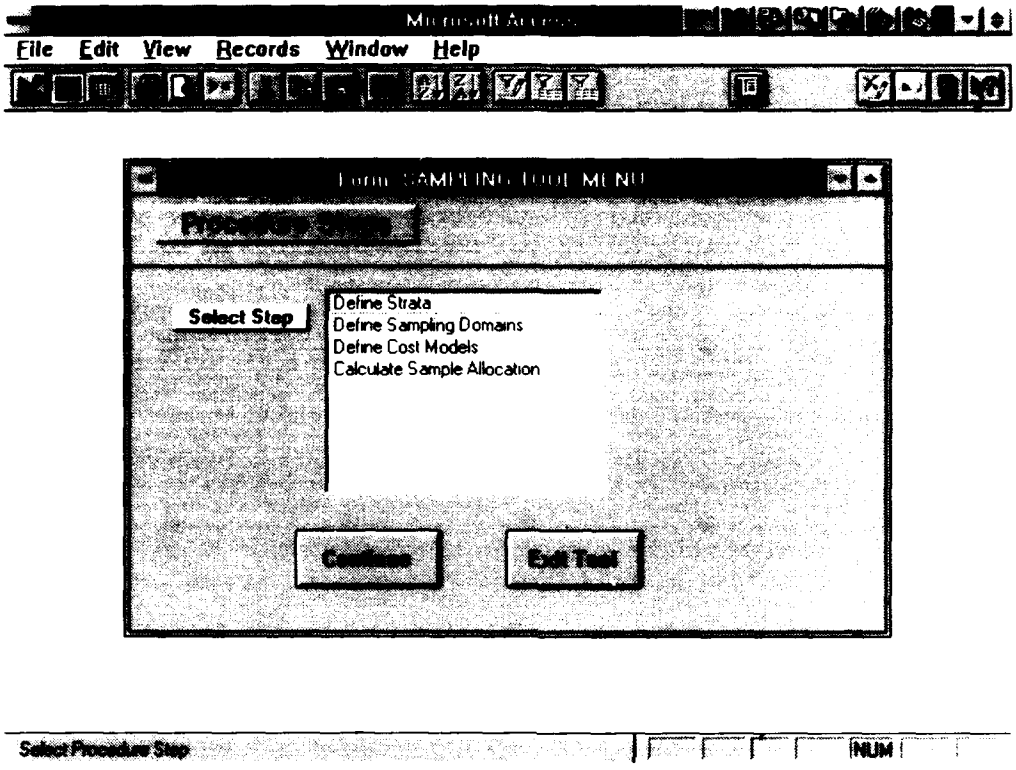


Once the Tool has been initialized the user may proceed with:

- defining the dimensions and levels of stratification,
- computing the stratum sizes,
- defining the reporting domains to be used in setting the precision requirements for the survey,
- setting the precision requirements,
- computing the domain within stratum sizes and marginal domain sizes,
- defining the cost model,
- setting the cost coefficients specified by the cost model
- setting the response rates to be used, and,
- computing the sample allocation.

Usually in the course of developing a sampling design several iterations are needed to achieve the desired results. Stratum definitions, domain definitions and precision constraints may be modified several times in response to the determined sample allocations. The Tool has been designed to facilitate making these types of changes during the course of the design development.

Sampling Tool Menu

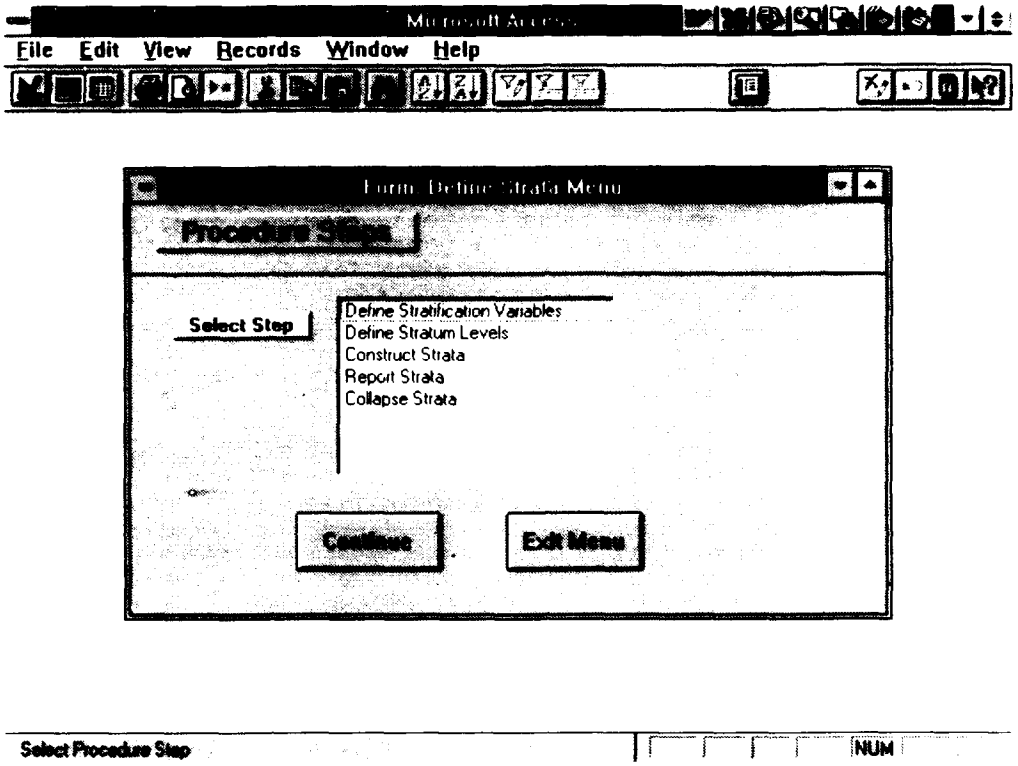


The *Sampling Tool Menu* is the first screen displayed when the sampling tool is opened after the identification screen on the previous page disappears. All steps used in calculating the sample allocation are invoked from this originating menu. To select an item, highlight it by clicking the line item with the mouse. Once the procedure choice is made, press the *CONTINUE* button. Another menu will appear listing the individual steps to be completed. To exit the sampling tool, press the *EXIT TOOL* button.

To close the *Sampling Tool Menu* without exiting Access, click on the *CONTROL-MENU BOX* on the upper left hand corner of the menu bar and press CLOSE.

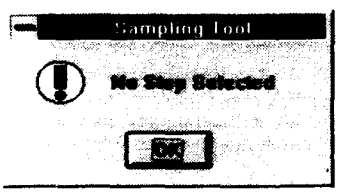
Menus, procedure step forms and procedure step reports are discussed in the remainder of this section. Some general rules apply across all three types of screens. First, the mouse pointer takes the form of an arrow on the screen. When the arrow is transformed into an hourglass, the system is processing a command such as opening a form or saving data to a table. No action should be taken with the mouse until the arrow has returned. Second, the system will beep when an error has occurred, when a decision is required to continue with a process, or when a lengthy process has completed. A message box or status bar message will indicate the appropriate action. These points are discussed throughout the manual.

Overview Of Tool Menus



All procedure steps that need to be completed to compute the sample allocation are invoked from a menu screen. In Access terminology, the Tool menus are forms. However, for consistency and clarity we will use the term menus to mean those screens which direct the invocation of task specific forms and the term forms to mean the screens which are used to complete a specific task.

All menus have the same basic structure consisting of a list box of procedure steps, a *CONTINUE* command button and an *EXIT MENU* command button. A procedure step is selected with a single click of a mouse on the relevant line item. Once a step has been selected, the *CONTINUE* button will highlight in red signifying that the Tool expects this button to be selected next. Clicking the *CONTINUE* command button will cause the form for the procedure step to invoke; this form is displayed overlaying the menu.



If the *CONTINUE* button is pressed prior to choosing a procedure step, a single button message box with a bright yellow exclamation mark will appear telling the user that a procedure step has not been selected. The menu is returned to it's original condition after the *OK* button has been pressed.

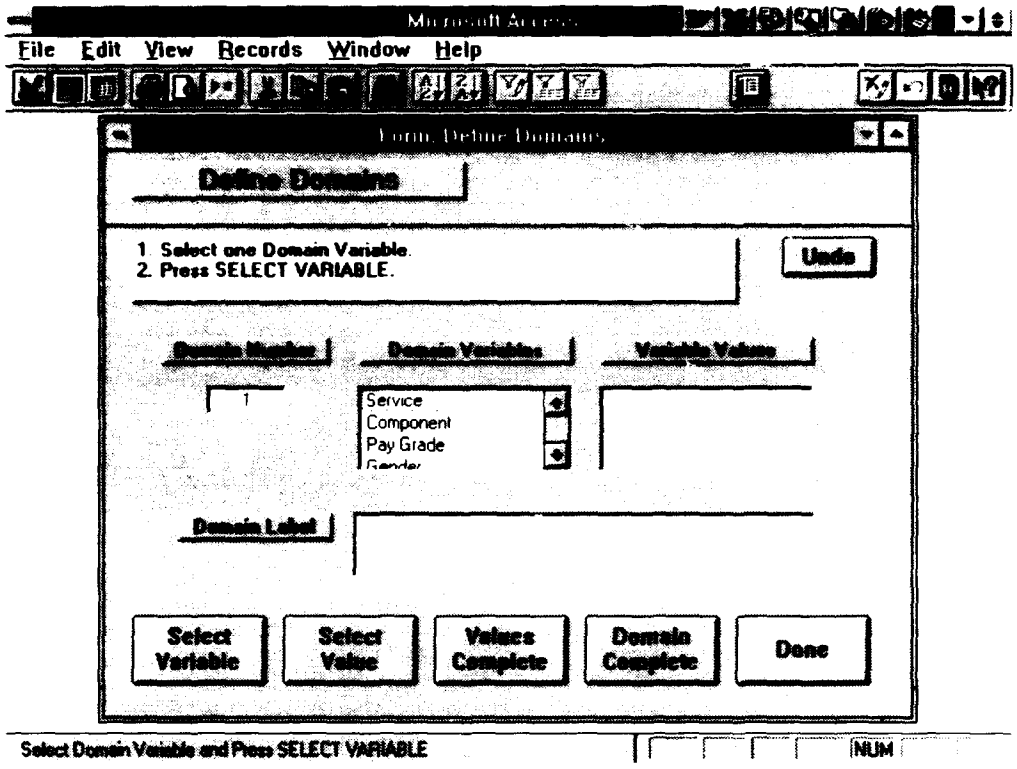
The tool assumes that the procedure steps will be completed in the order in which they are listed in the menu. That is, forms activated at a given step may require information obtained in a previous step. Since information is retained in tables as part of the database, the steps

need not be completed within the same sampling tool session. Message boxes, similar to the one shown above, will appear if the table information required by a process step does not yet appear in the database.

However, changes made in the course of the total design development will necessitate repeating some of the procedure steps. If a step is repeated, then the tables generated in steps following the repeated step will exist, but will contain information that is no longer consistent with the changes made in the repeated step. Consistency is re-established by repeating these steps also.

The *EXIT MENU* button closes the menu. If the menu in question is the *Sampling Tool Menu*, then this command button additionally closes the database and exits Access. On all other menus the *EXIT MENU* button returns control to the *Sampling Tool Menu*.

Overview Of Tool Forms

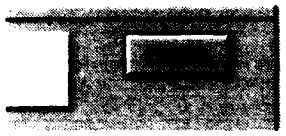


Procedure steps, such as the defining sampling domains step depicted above are completed using forms invoked from a particular menu. All forms have a descriptive title (*Define Domains*) to indicate which task is to be completed. Instructions and general information are usually provided in two places, a yellow box near the title and the status bar at the bottom left corner of the Microsoft Access screen. The status bar will contain either instructions or a blue meter that displays the percentage of some action that has been completed, such as the percentage of records that have been processed.

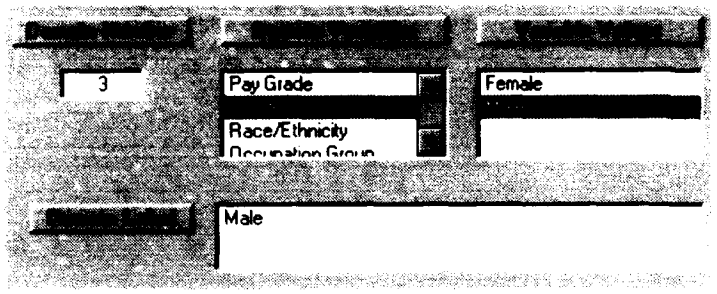
Objects associated with the next action to be taken in the form, such as a command button or a list box, are highlighted in red. In the form above, the user is required to select a variable to be used in the declaration of Domain #1. Once a variable has been selected, the red highlight is transferred from the label *DOMAIN VARIABLES* to the words *SELECT VARIABLE* on the *SELECT VARIABLE* command button at the bottom of the form.

Most forms will have text boxes, list boxes and command buttons. Text boxes are the white rectangles on a form in which one or more numeric values are to be read or entered. On the form above, *DOMAIN NUMBER* is a read-only text box. List boxes, similar to those used in menus, will contain information from which a selection can be made (*DOMAIN VARIABLES*) or just viewed (*DOMAIN LABEL*). Command buttons, such as *SELECT VARIABLE* and *DONE*, cause the form to process defined information and invoke an action.

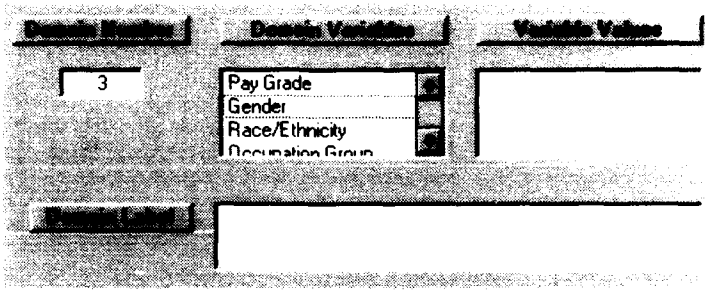
The Tool is programmed so that the values entered in text boxes and the selections made from list boxes may be changed prior to pressing a command button. For example, a selection can be made from the *DOMAIN VARIABLES* list box and changed prior to pressing the *SELECT VARIABLE* command button. Once this button is pressed, value codes for the selected variable are processed and displayed in the *VARIABLE VALUE* list box.



Most but not all forms will have an *UNDO* button. When this button is pressed, information for the current action, such as defining a domain, will be erased and the form reset. It is important to note that this button will not delete any data that had been previously saved.

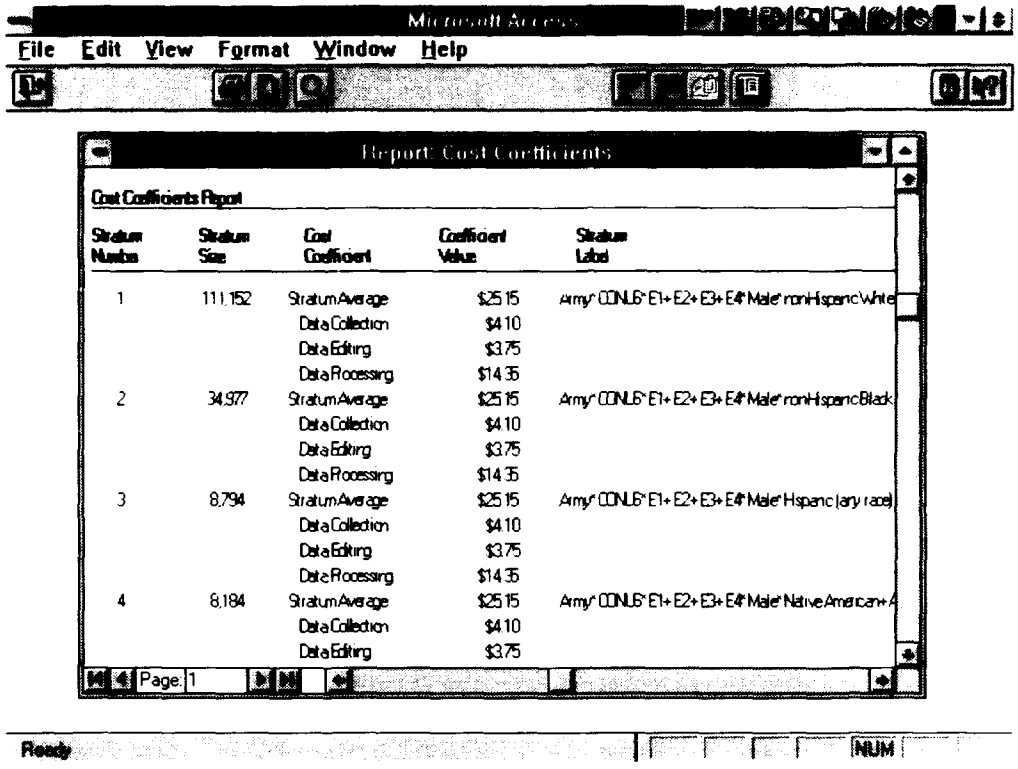


In this example domain 3 has been defined as male but should be female. If *DOMAIN COMPLETE* has not been pressed, then pressing *UNDO* will clear the *DOMAIN LABEL* and re-set *DOMAIN NUMBER* to 3.



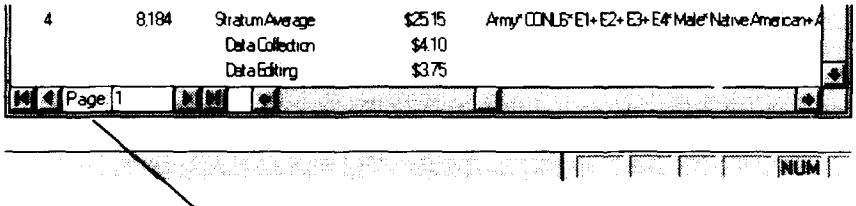
Because its behavior differs on different forms, the *UNDO* command button is discussed in what follows with reference to the relevant forms.

Overview Of Tool Reports



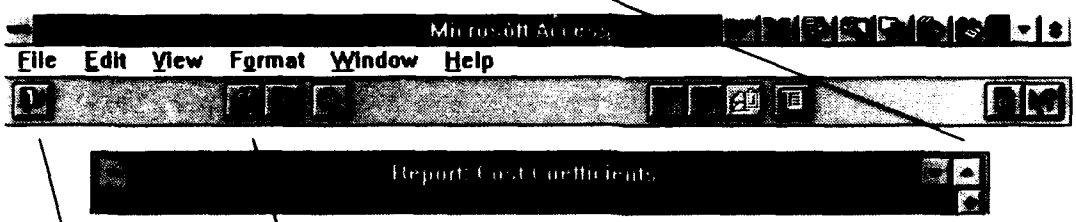
Reports are used to display information produced within the procedure steps. For example, the above report (*Cost Coefficients Report*) provides a listing of the coefficients used in the cost model for each sampling stratum. Other information, such as *STRATUM SIZES* and *STRATUM LABEL*, is also listed. If this form is invoked prior to the *STRATUM AVERAGE* information being defined, then zeros are shown in place of the stratum average cost coefficients. Missing information is also depicted as blank entries in some reports.

Note that any previously reported information is lost with the invocation of a new report.



Once the report is displayed on the screen the document may be paged through by using the page selector in the bottom left corner.

The screen may be expanded to its full size by pressing the up arrow in the upper right corner of the report.



Provided that a printer has been connected prior to invocation of the Tool, the second button on the toolbar (a printer) will cause a hard copy of the report to be produced.

The left-most button on the formatting toolbar (an open file folder with an arrow) will close the report and return control to the calling menu.

Constructing Strata

Stratification provides a mechanism for controlling the distribution of the sample with respect to characteristics of interest to the investigation. Usually the characteristics chosen as stratification variables are associated with data collection costs and with the need to have predetermined representation in the sample of specified subpopulations. The subpopulations can be defined for example by factors such as geography, demographics, and differential response rates or other measurement problems. To be useful as a stratification variable the identified characteristic must be known for essentially each element in the sampling frame. Elements for which one or more characteristics are unknown can be grouped together in an "unknown" or "other" stratum, but the size of this group must be relatively small if the stratified design is to be truly effective.

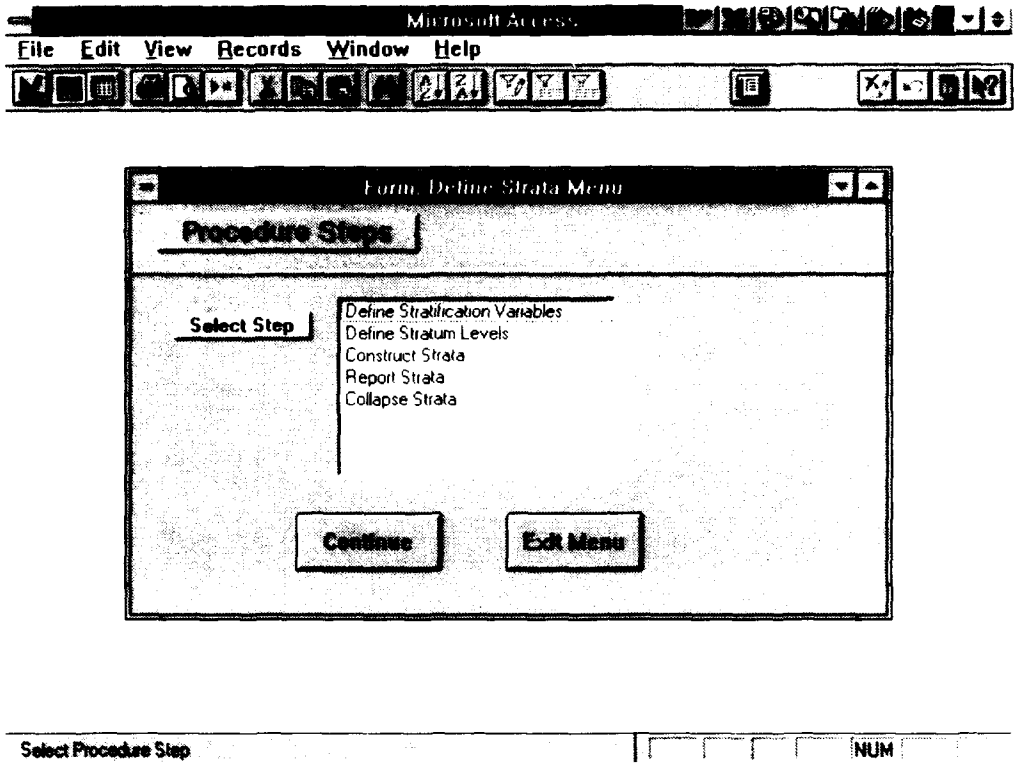
As the term is usually used, stratification partitions the frame in the mathematical sense. That is, each element in the frame can only belong to a single stratum. Further, in the aggregate the strata completely account for the entire population of interest. If this is not the case, then non-coverage biases arise in association with the incomplete frame. The magnitude and direction of the biases depend on the number of population units excluded from the sample and on the differences between the excluded and included units.

Most often samples are independently selected in different strata. A sample size must be specified for each stratum. Different sampling designs and estimation procedures and/or different data collection designs and measurement procedures may be used in different strata, though this is an infrequently used method. As implemented by the Tool, the user first identifies those variables in the "Source Data" table that will be used to define the strata. In the Tool terminology, the stratification variables define the dimensions of stratification. Next, the user identifies which variable values or codes are to be used in constructing the strata. The variable values define the levels of stratification in each dimension; a level may be defined by a single variable value or by groups of variable values.

As the Tool is currently written, strata are initially constructed by cross-classifying each of the defined dimensions. In general a complete cross-classification is likely to result in some very small, perhaps totally empty, strata. Depending on the total sample size, very small strata may be too small to make an important contribution to controlling the distribution of the sample, and may even contribute to some difficulty in initially selecting the sample and later in computing variances. Certainly empty strata need to be eliminated from further consideration. However, despite these difficulties the cross-classification provides a convenient mechanism for indexing the strata within the Tool tables. This aids in keeping track of the cumulative changes that might be made in the dimensions and levels of stratification in the course of the design. On the other hand, this approach necessitates programming the additional capability to collapse some strata into others, to increase the size of the resulting strata and for other purposes.

The forms described in this section provide the user interface for identifying the variables and variable values to be used in stratifying the sampling frame, for constructing the strata, and for subsequently collapsing the strata should this be necessary. The Tool report for this step lists the stratum identifiers and stratum sizes and identifies the variables and variable values used to define the strata.

Define Strata Menu

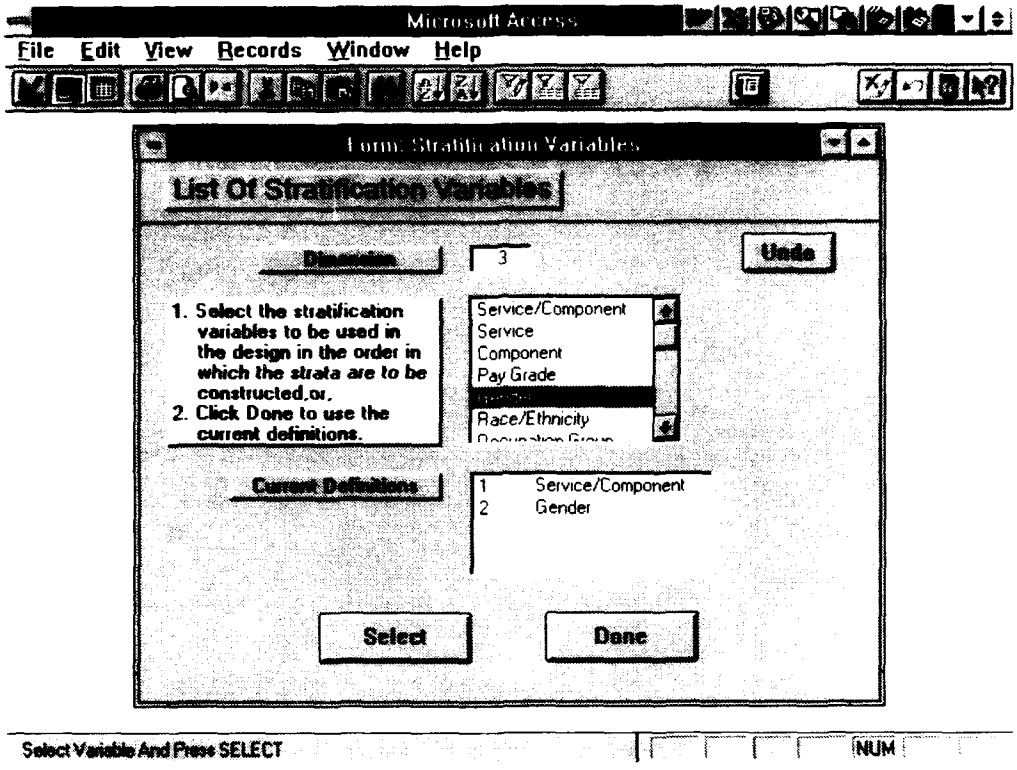


The second menu in the sample allocation sequence is the *Define Strata Menu*. All of the procedure steps associated with the construction of the sampling strata are completed from this menu. The menu items deal with:

- defining the dimensions of stratification,
- defining the levels of stratification,
- assigning the stratum identification numbers and computing the stratum sizes,
- collapsing selected levels of stratification,
- reporting the stratum definitions and stratum sizes.

As with the originating menu (*Sampling Tool Menu*), a form is invoked by clicking the corresponding line item with a mouse and pressing *CONTINUE*. The *EXIT MENU* button is pressed to close the menu and return control to the originating menu.

Stratification Variables Form

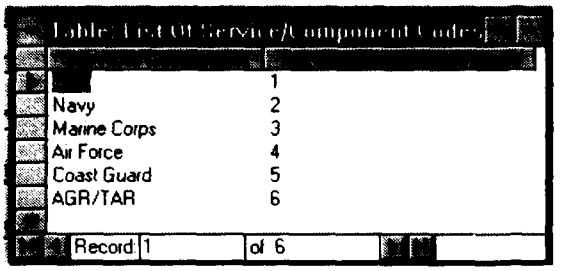


The *Stratification Variables Form* is the first of two forms used to the create sampling strata. This form enables the declaration of the stratification variables or dimensions. All variables listed in the "Source Data" files, except the variable *COUNT* (see "Source Data" discussion, pg. 2), are included in the variable list box and are candidates for use as stratification variables.

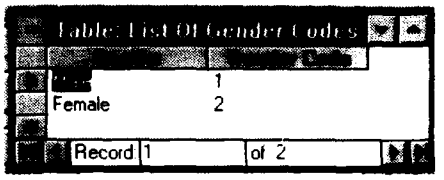
If dimensions of stratification were defined in a previous session, then their labels are provided in *CURRENT DEFINITIONS* list box. Otherwise, the list box will contain "none." New dimensions are chosen by clicking a variable name and then pressing *SELECT*. As the dimensions are selected, the *DIMENSION* number is incremented and the variable name selected is added to the *CURRENT DEFINITIONS* list box.

The *DONE* button is pressed after all dimensions of stratification have been defined or if the definitions defined previously are to be used.

In this example above two dimensions of stratification have been defined, Service/Component and Gender.

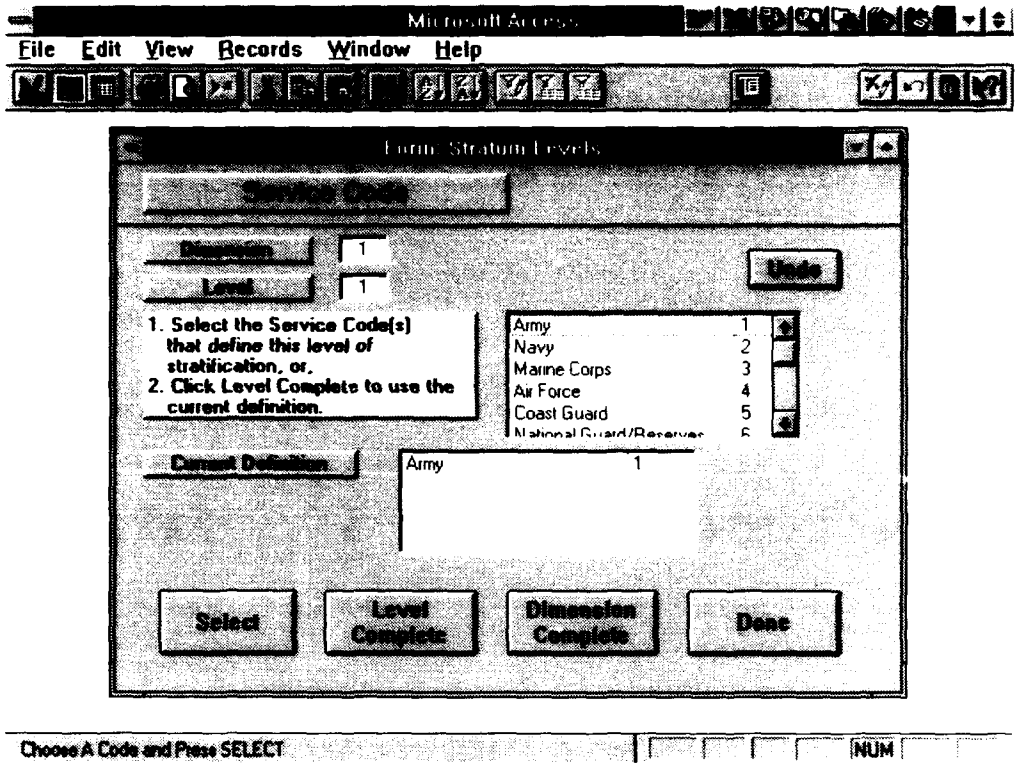


The relevant code tables contain six Service/Component codes and two Gender codes.



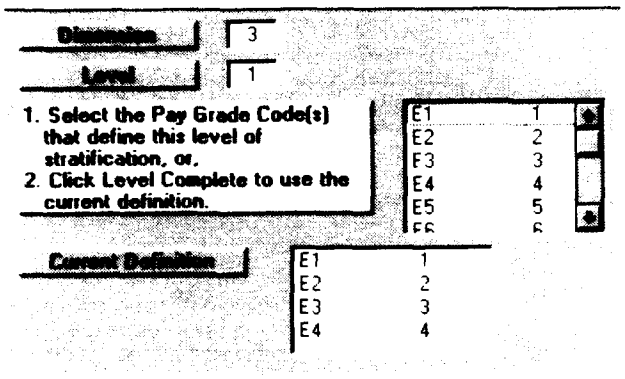
Thus at this point two dimensions of stratification have been defined, one potentially at six levels and the other potentially at two levels. If no changes are made, then 6 x 2 = 12 strata can be defined. However, not all of the codes listed in the code tables need to be used to construct strata. Further, several codes may be combined to form a single level of stratification.

Stratum Levels Form



The *Stratum Levels Form* is the second of two forms used to create sampling strata. The levels of stratification in each dimension are defined in this form. As in the *Stratification Variables Form*, any previously defined levels created for the relevant dimensions are displayed in *CURRENT DEFINITION* list box. If there are no previous definitions, then default levels listed in the corresponding code tables are provided in the list box.

The previous level definitions or the default levels are used simply by pressing *LEVEL COMPLETE* command button. Alternately, new levels are defined by clicking the variable values or codes listed and pressing *SELECT*. As each value is chosen and *SELECT* pressed, the value label and code is listed in the *CURRENT DEFINITION* list box.

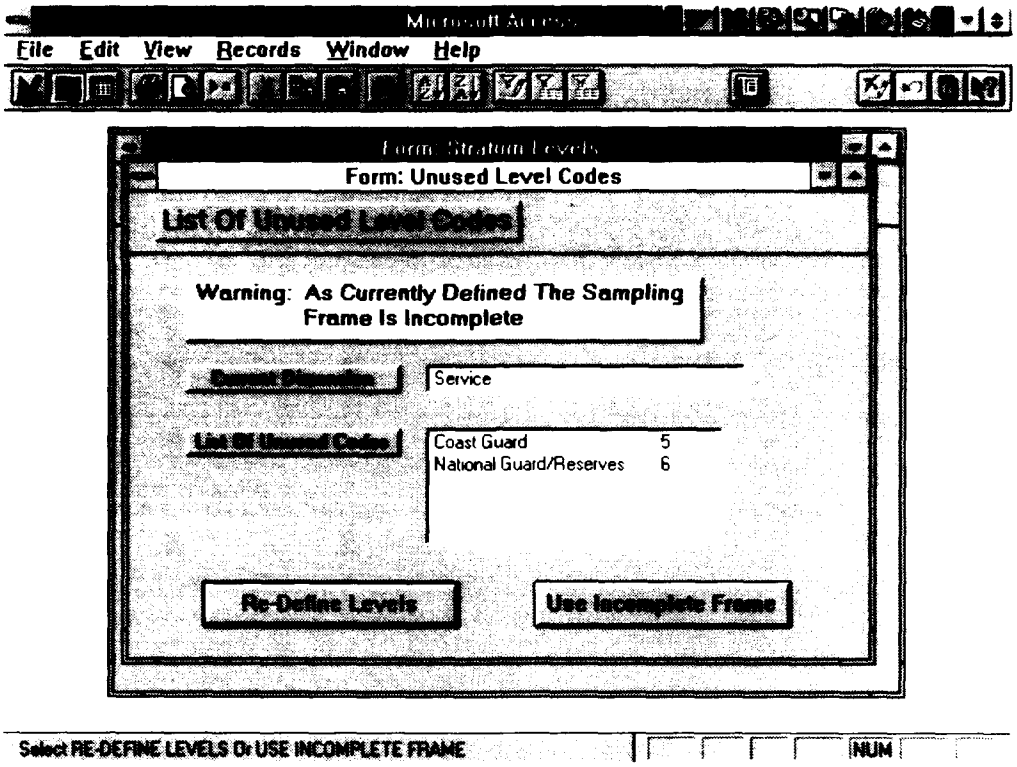


Multiple values or codes can be used to define a single level. In this example the Pay Grades E1 through E4 have been combined to define level 1 within dimension 3, Pay Grade.

Press *LEVEL COMPLETE* to signal that all of the codes needed to define the level of stratification have been identified. Control passes back to the code list. Press *DIMENSION COMPLETE* after all levels have been defined for the dimension in question.

If an incorrect definition is created for a level, *UNDO* clears the definitions for all of the levels within the current dimension and resets *LEVEL* to one. *DONE* will save all dimension level information to the table "New Stratum Key" and close the form. Control returns to the *Define Strata Menu*.

Unused Level Codes Form



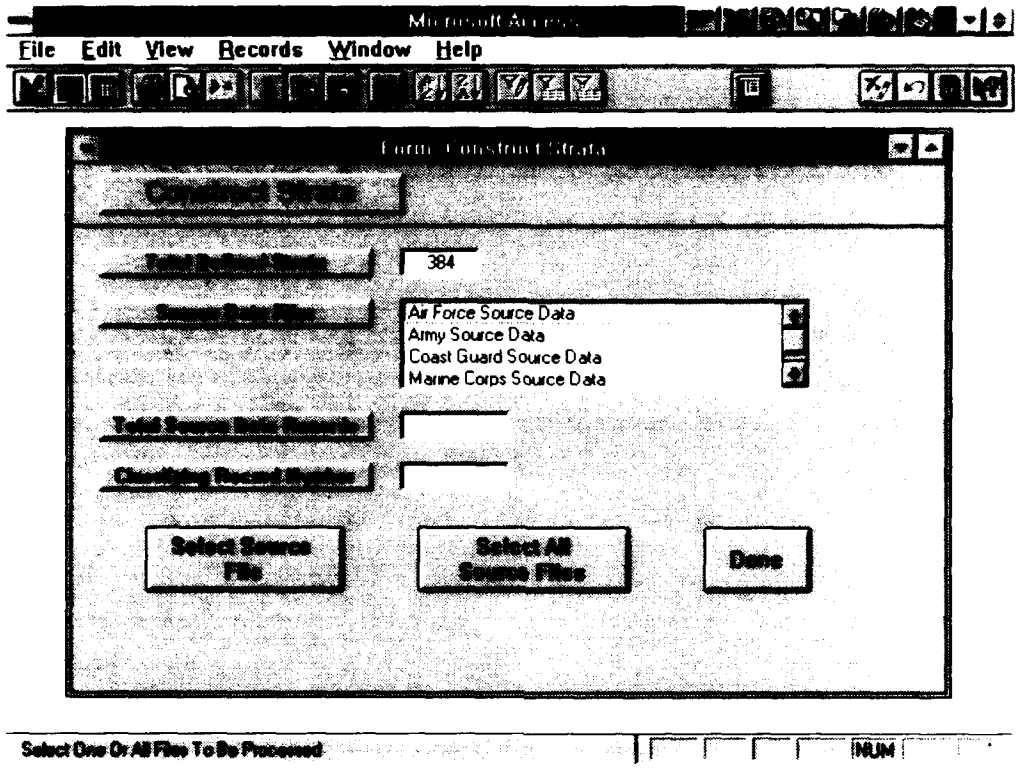
If one or more of the codes provided in a code table are not assigned to a level of stratification in the *Stratum Levels Form*, the *Unused Level Codes Form* will activate. The purpose of this form is to call to the user's attention the fact that the sampling frame may be incomplete. The codes in the code table that have not been used to define a level of stratification are listed in alphabetical order in the *LIST OF UNUSED CODES* list box for the *CURRENT DIMENSION*.

In the above example the sampling frame is being stratified by Service. Codes identifying the Coast Guard and National Guard/Reserves are provided in the relevant code table but have not been used to define levels of Service. The omission might be an oversight or intended given the scope of the survey in question.

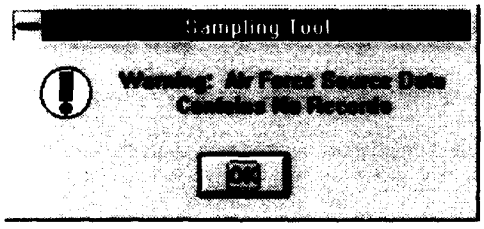
Press *RE-DEFINE LEVELS* if omission of the Coast Guard and National Guard/Reserves is an oversight. Control returns to the *Stratum Levels Form* with *LEVELS* reset to one and *CURRENT DEFINITIONS* set to "none" for the *CURRENT DIMENSION*.

Press *USE INCOMPLETE FRAME* if the Coast Guard and National Guard/Reserves are to be intentionally omitted. Control returns to *Stratum Levels* with *DIMENSION* appropriately incremented and *LEVEL* set to one.

Construct Strata Form



Once the dimensions and levels of stratification are specified, the *Construct Strata Form* is used to calculate the sizes of the defined strata. The total number of strata computed by crossing all of the dimensions is shown in the *TOTAL DEFINED STRATA* text box. All of the available source data files (those tables with "Source Data" at the end of the name) are displayed in the *SOURCE DATA FILES* list box. From this list the data sets to be used in the calculation of the stratum counts are specified.

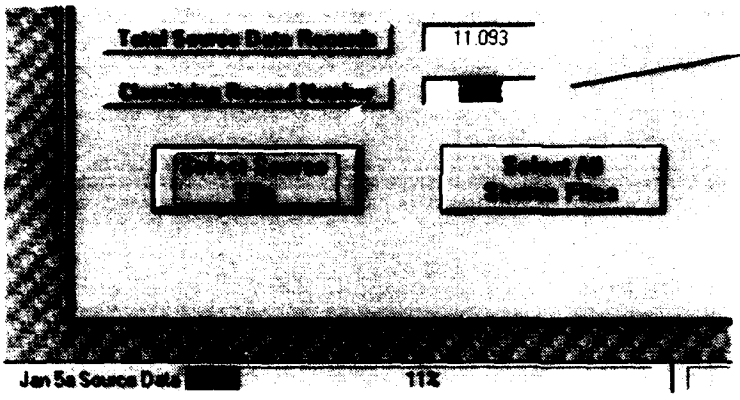


If a data file exists but contains no records, then a single-button message box appears with this information. Only exiting the sampling tool and importing the table will repair this problem.

If not all of the source data files are to be utilized, select an appropriate file by clicking the corresponding line item in the list box and pressing the *SELECT SOURCE FILE* command button. Additional files are chosen after each data set is processed.

The *SELECT ALL SOURCE FILES* command button will cause the program to process all valid source data files in the database without having to select each line item. The total

number of records in the selected "Source Data" table is displayed in *TOTAL SOURCE DATA RECORDS* once a command button is pressed.

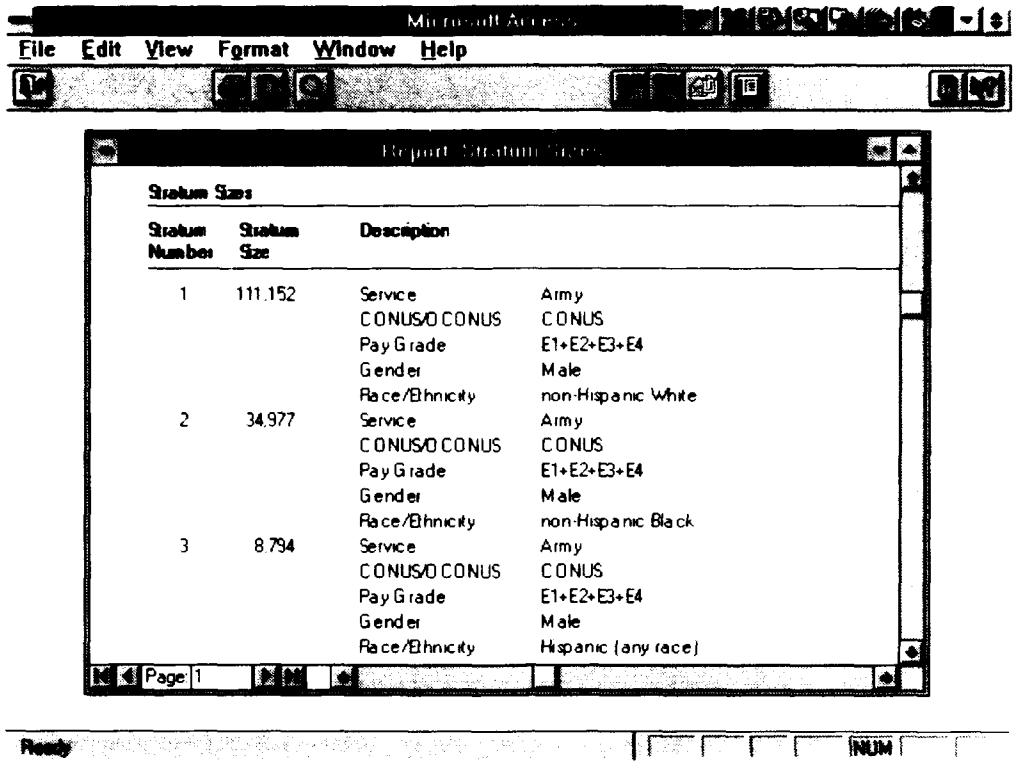


As records are processed, the overall cumulative stratum count is exhibited in the *CLASSIFYING RECORD NUMBER* text box.

The percentage of records processed is shown on the blue meter positioned in the lower left portion of the screen. Stratum sizes are stored in the table "Stratum Sizes."

Press *DONE* once all relevant files have been processed.

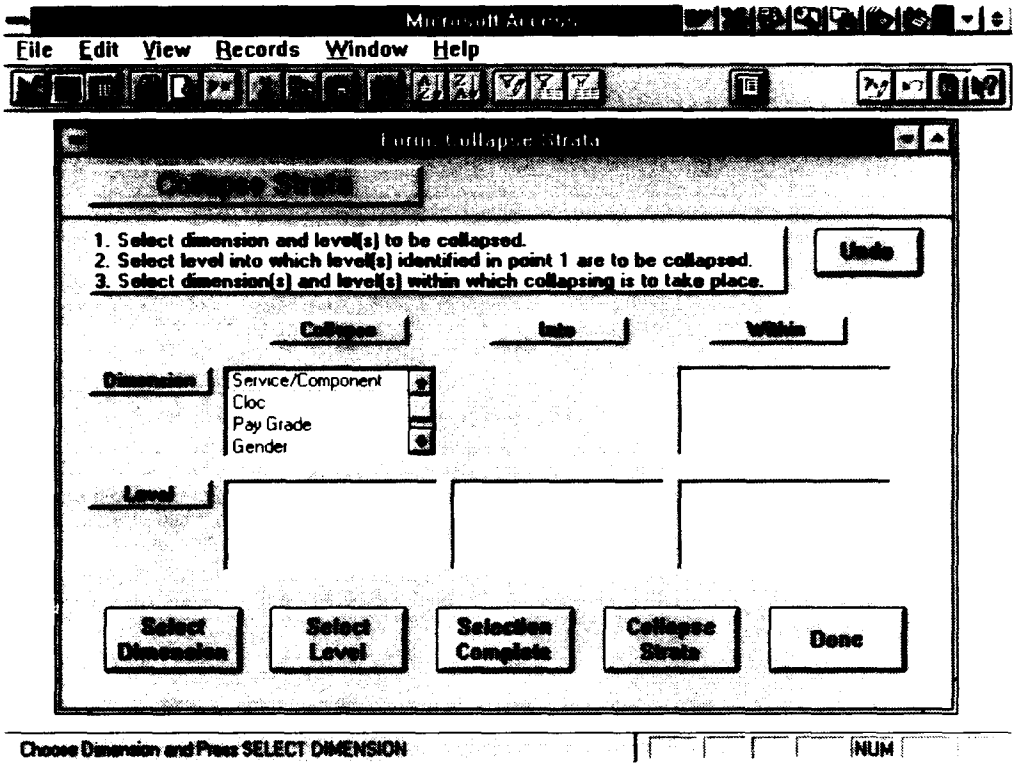
Stratum Sizes Report



The first in the series of sampling tool reports is the *Stratum Sizes Report*. The purpose of this report is to allow the identification of strata containing a "small" number of subjects (*STRATUM SIZE*). An absolute minimum of two observations in any stratum is required for the calculation of variance estimates. However, for a stratum-level estimate to have any reasonable power, each stratum should have quite a few more than two observations. Strata which are judged to be too small are combined using the *Collapse Strata Form* described next. Obtaining a printout of the *Stratum Sizes Report* is strongly suggested for reference while using the *Collapse Strata Form*.

The *DESCRIPTION* column on the report lists the dimension of stratification, such as Service, followed by a level within the dimension, such as Army. When the report is invoked, a blue meter is provided in the bottom left corner of the screen to display the percent of records processed in the course of generating the report.

Collapse Strata Form



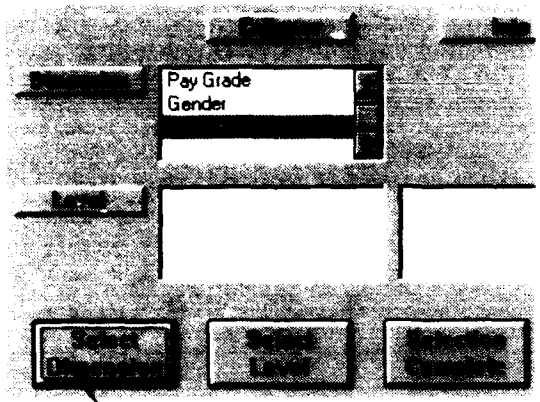
The *Collapse Strata Form* is used to collapse strata that are judged to be too small or that may need to be collapsed for other reasons. The form opens with the dimensions of stratification displayed in the *DIMENSION* list box.

		Race/Ethnicity	non-Hispanic White
246	228	Service/Component	AGR/TAR
		Cloc	CONUS
		Pay Grade	E1+ E2+ E3+ E4
		Gender	Female
		Race/Ethnicity	non-Hispanic Black
247	84	Service/Component	AGR/TAR
		Cloc	CONUS
		Pay Grade	E1+ E2+ E3+ E4
		Gender	Female
		Race/Ethnicity	Hispanic (any race)
248	32	Service/Component	AGR/TAR
		Cloc	CONUS
		Pay Grade	E1+ E2+ E3+ E4
		Gender	Female
		Race/Ethnicity	Native American+ Asian & Pacific Islander+ Other

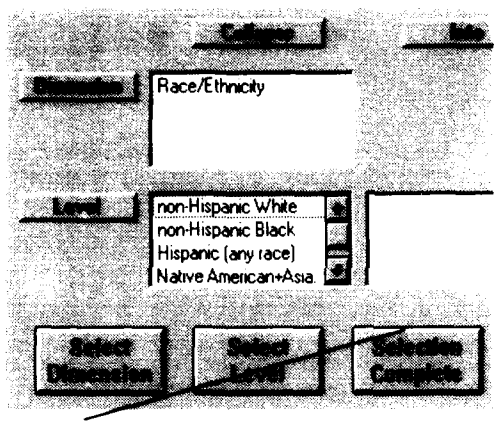
The procedures used in this form are explained using the following example. Suppose that strata have been constructed with the five dimensions shown in this excerpt from *Stratum Sizes Report*.

The decision is made to collapse stratum numbers 248 and 247 into 246 to form a new level.

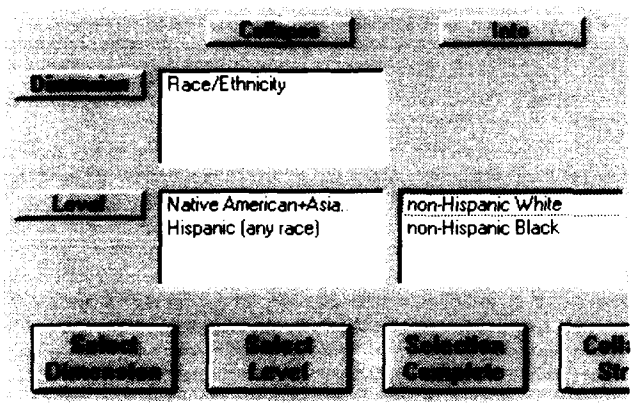
The new level combines non-Hispanic Black, Hispanic (any race), and Native American + Asian & Pacific Islander + Other into a single level within the Race/Ethnicity dimension. For the example collapsing is to be completed only for female enlisted grades E1+E2+E3+E4 within the CONUS location of AGR/TARs.



Select the dimension to be collapsed, namely Race/Ethnicity, by clicking its label in the *COLLAPSE DIMENSION* list box and then pressing the *SELECT DIMENSION* command button. All remaining dimensions are cleared from the list. A list of levels for Race is provided in the *COLLAPSE LEVEL* list box.

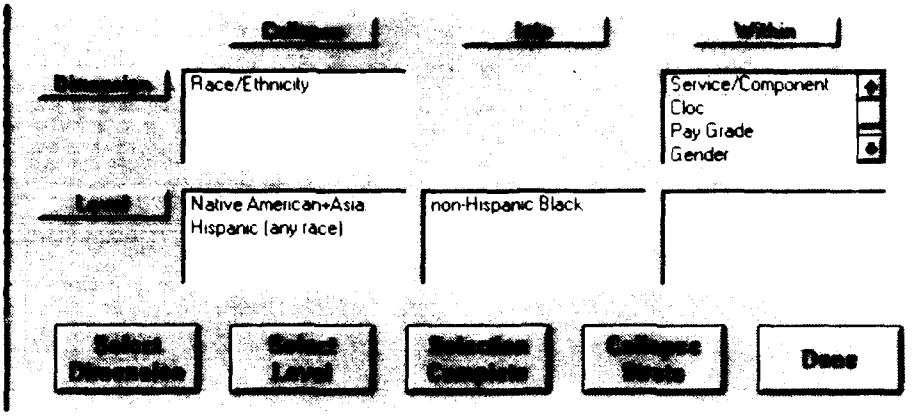


Select in turn the levels to be collapsed, Native American + Asian & Pacific Islander + Other, then Hispanic (any race). First highlight a label in the *COLLAPSE LEVEL* list box and then press *SELECT LEVEL*. After both selections are complete, press *SELECTION COMPLETE*.

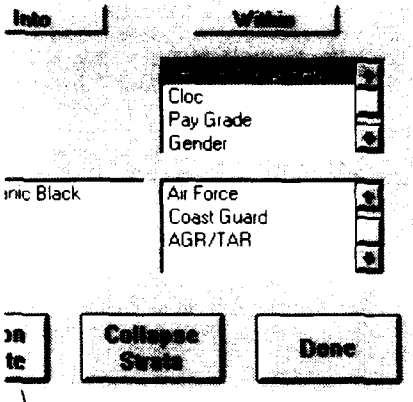


Next select the level within the Race/Ethnicity dimension into which the levels selected above are to be collapsed, that is non-Hispanic Black. The list box *INTO LEVEL* lists the available choices. Select the "into" level by first clicking its label and then pressing the *SELECT LEVEL* command button.

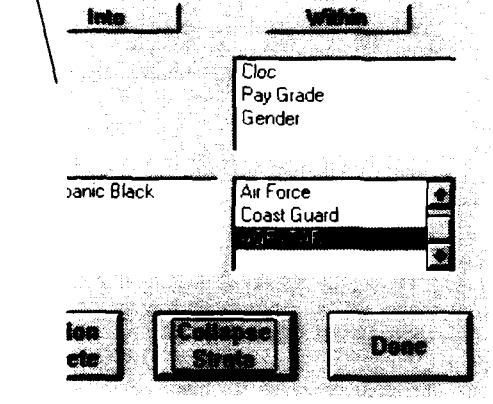
The information in the list boxes shows the selections made to this point. The complete screen for the example is shown below.



One step remains, namely identifying the dimensions and levels of stratification within which the Race/Ethnicity collapsing is to take place. For the example, the collapsing is to be completed for AGR/TAR female enlisted grades E1 through E4 in CONUS. The complete list of "within" dimensions, all of them except the Race/Ethnicity dimension, is provided in the *WITHIN DIMENSION* list box.



Select each *WITHIN DIMENSION* in turn. First click the dimension label in the *WITHIN DIMENSION* list box and then the *SELECT DIMENSION* command button. The list of available levels in the selected dimension is shown in the *WITHIN LEVEL* list box.



Select the level or levels by clicking the level label in the *WITHIN LEVEL* list box and pressing the *SELECT LEVEL* command button. When a dimension level is selected, the dimension label is removed from *WITHIN DIMENSION* once control returns to the list box. When all of the necessary levels have been selected, press *SELECTION COMPLETE*.

This process is continued until either the *SELECTION COMPLETE* command button is pressed or all of the dimensions listed in the *WITHIN DIMENSION* list box have been exhausted. The order in which the "within"

dimensions are selected is not important but following the order of the *Stratum Sizes Report* is helpful.

Once the "within" dimensions and levels have been selected and *SELECTION COMPLETE* has been pressed, pressing the command button *COLLAPSE STRATA* causes the collapsing action to be carried out and the size of the newly created stratum to be computed. Strata that are collapsed have a -1 in place of the stratum size in the *New Stratum Sizes* table. The stratum number into which each stratum has been collapsed is provided in the *COLLAPSE STRATUM* field of the "New Stratum Sizes" table. The collapsed strata are eliminated in the *Stratum Sizes Report*.

The form re-initializes and control returns to the *COLLAPSE DIMENSION* list box after the collapsed stratum sizes have been computed. Note that at any point in this form prior to pressing the *COLLAPSE STRATA* command button, the *UNDO* button may be pressed to erase the current set of specifications and re-start the collapsing procedure at the *COLLAPSE DIMENSION* list box.

Another set of strata may be processed or the form closed by pressing the *DONE* command button.

Checking the stratum sizes after collapsing is recommended to ensure that the intended collapsing scheme was accurately implemented. If this procedure needs to be completed from the beginning with an uncollapsed table, simply delete the "New Stratum Sizes" table. Deleting this table altogether causes the procedure to re-start using the table "Stratum Sizes" which contains the original set of strata created by cross-classifying all of the dimensions.

If, instead of AGR/TAR used in the above example, the Race/Ethnicity dimension was to be collapsed in all of the Services for female enlisted grades E1 through E4 in CONUS, then select all *WITHIN DIMENSIONs* and appropriate *LEVELs* except Service.

The Race/Ethnicity dimension cannot be collapsed across all other dimensions within one iteration of the procedure; at least one *WITHIN DIMENSION* and *LEVEL* must be specified. If this is the desired result, two options exist. The first option is to invoke the *Stratum Levels Form* and re-defined the Race/Ethnicity levels. Stratum sizes (*Construct Strata Form*) will need to be computed again. The second option is to collapse the Race/Ethnicity dimension across all levels of a particular dimension, say Males and Females within the dimension Gender.

Imposing Precision Requirements

The sample size and allocation are determined in general in response to precision requirements developed as an integral part of the statement of the objectives of the survey. The precision requirements take the form of inequality variance constraints imposed on identified key parameter estimates. The Tool is based on the premise that the parameters of interest are defined in terms of population proportions. This premise negates the necessity of knowing the values of the relevant population variances which would otherwise be required to compute the sampling variances of the estimates. The (binomial) population variances are coincidentally specified in specifying the values of the proportions.

Several steps are involved in specifying the precision requirements. Most usually the key parameter estimates take the form of the proportion of one or more reporting domains that possess some attribute of central interest to the survey. The implied steps involve:

- defining the domains of interest, and
- specifying the proportion of domain members to be estimated.

Examples of possible reporting domains include:

- all women in the military,
- all women in the Navy,
- all female field grade officers in the Navy, and
- all female field grade officers in the Navy who are belong to a racial minority.

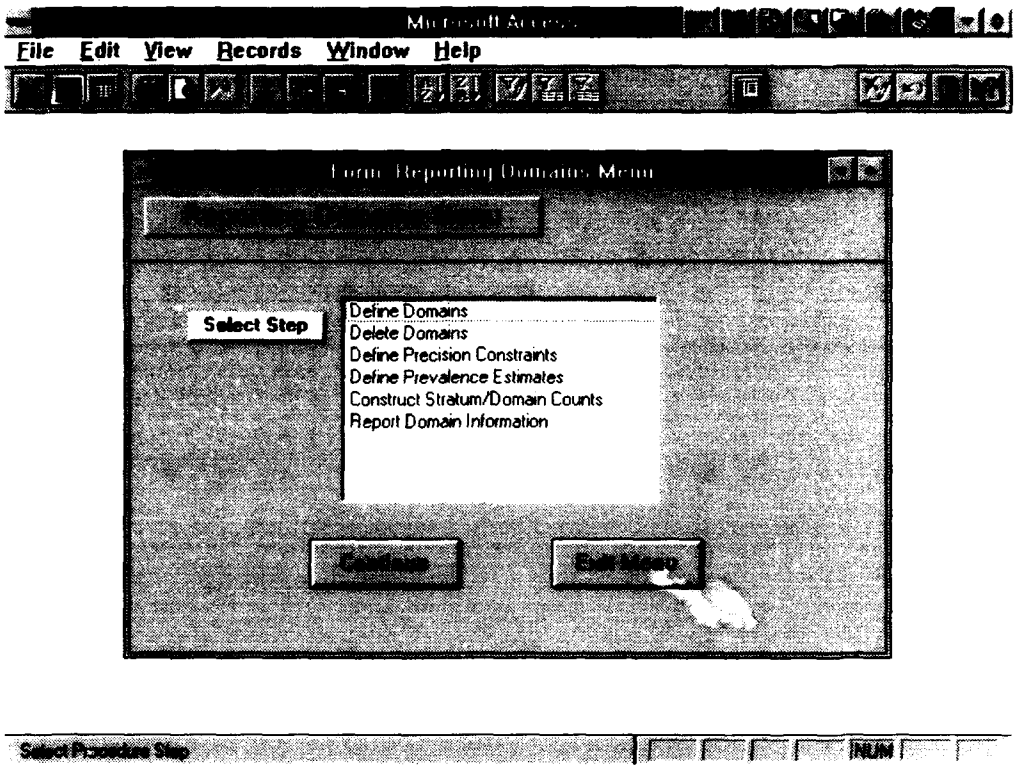
The proportions might be the relative numbers of domain members who have experienced one or more incidents of unwanted gender-based attention during a specified time period. The proportions are referred to as prevalence estimates on the forms comprising this section and in what follows.

Given the domain definitions and the prevalence estimates for each domain, the next step involves specifying the maximum values of the sampling variances to be associated with the corresponding sample estimates. In general, the variance specifications can take a variety of forms depending on the preferences of the investigator. For example, rather than specifying the variances per se, different investigators might choose to use relative standard errors, confidence intervals, one- or two-tailed tolerance intervals, or, the size and power to be associated with formal tests of hypotheses. This version of the Tool requires that the variance specifications take the form of the maximum confidence interval half-widths to be associated with the prevalence estimates.

The process of imposing precision requirements tends to be an iterative process. Usually the originally imposed requirements turn out to yield total sample sizes that are in excess of budget realities. As a consequence the investigator is required to delete some domains, redefine others, modify the prevalence estimates, and change the variance constraints in a

search for a configuration that will meet the objectives of the study and also the cost constraints. The Tool has been developed to facilitate this type of development.

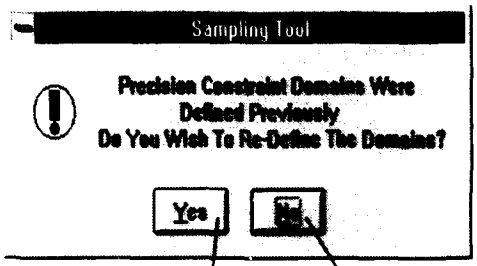
Reporting Domains Menu



The second menu in the sampling tool is the *Reporting Domains Menu*. All steps associated with the development of reporting domains are completed from this menu. The *CONTINUE* button will invoke any of the listed items. The *EXIT MENU* button is pressed to close the menu and return control to the originating menu.

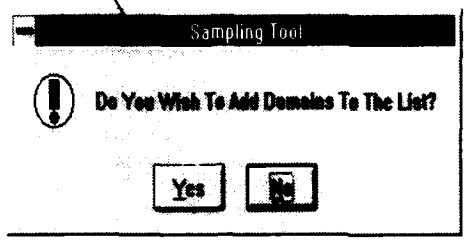
A discussion of the list item *DEFINE DOMAINS* is necessary at this point. If reporting domains were not created in a previous session, then *CONTINUE* will result in the form *Define Domains* being invoked. However, if at least one domain has already been created, then a series of two-button message boxes appears for determining the following courses of action.

- Re-defining all of the previously defined domains.
- Re-defining some of the previously defined domains.
- Adding new domain definitions to the current list.



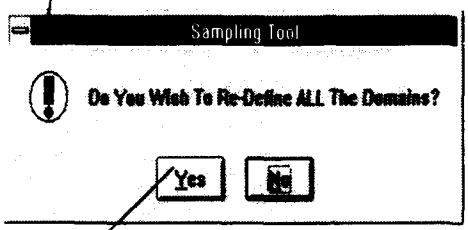
Press *YES* to re-define some or all of the currently defined domains.

Press *NO* to add new domains to the current list.



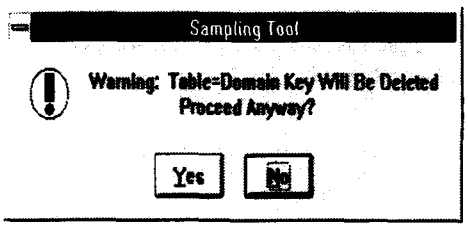
Press *YES* to add new domains to the current list.

Press *NO* to return to the menu without any changes.



Press *YES* to delete all currently defined domains and start from the beginning.

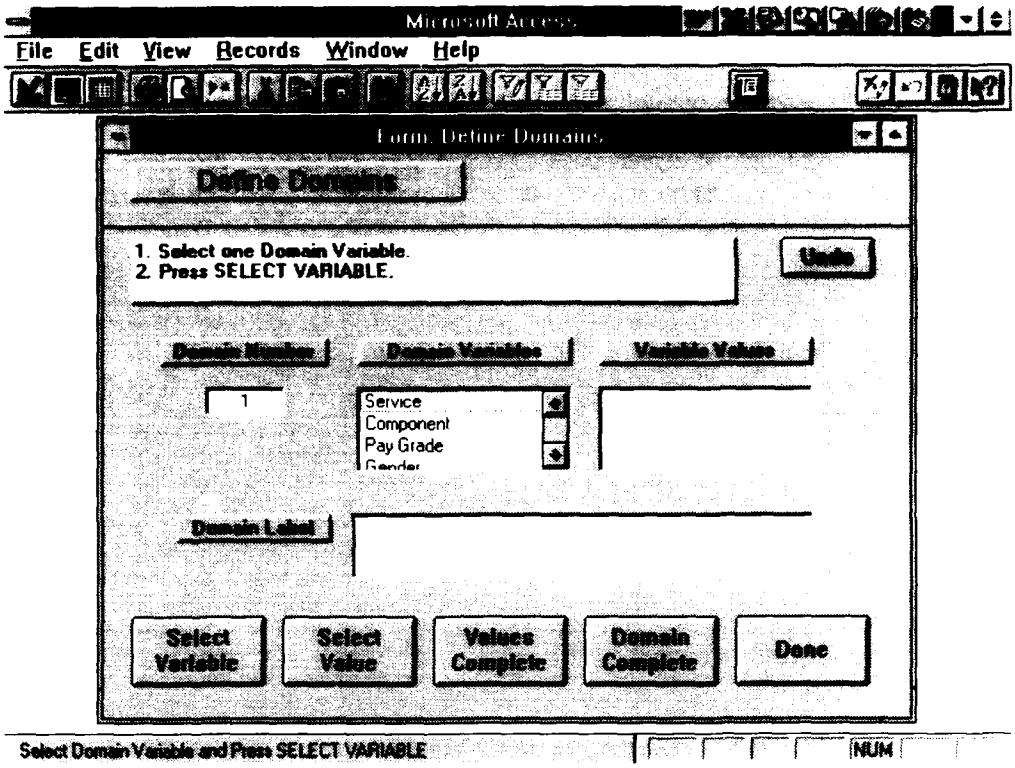
Press *NO* to re-define one or more of the currently defined domains.



Press *YES* to confirm the intent to delete all currently defined domains and start from the beginning.

Press *NO* to return to the menu without making any changes.

Define Domains Form



The form *Define Domains* is used to create reporting domains. Select the first domain variable and press *SELECT VARIABLE*. As in other forms the values associated with the selected variable are then listed in the *VARIABLE VALUES* list box.

Once a variable value is selected and *SELECT VALUE* pressed, the name appears in the *DOMAIN LABEL* list box and control returns to the value list. Selection continues until all relevant values for the specified domain variable have been chosen (ex. E1-E4 for Pay Grade).

VALUES COMPLETE will return control to the *DOMAIN VARIABLES* list where the next variable associated with the domain may be chosen. Values for subsequent variables are listed in *VARIABLE VALUES* and the process continues until the domain has been fully defined.

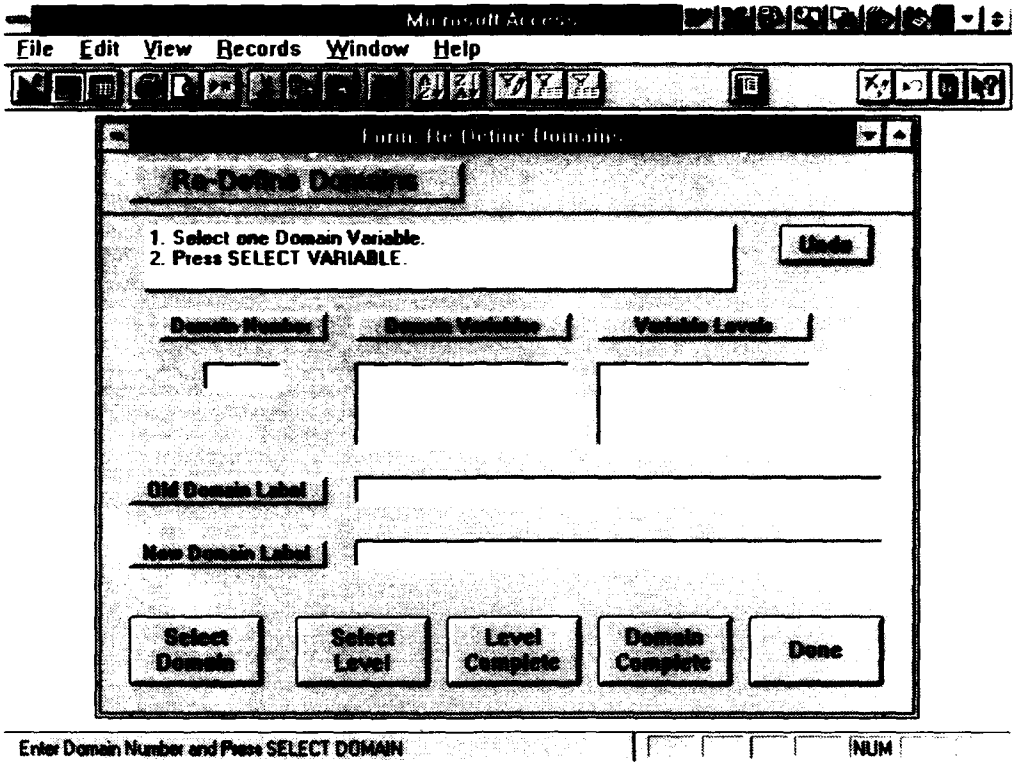
Clicking the *DOMAIN COMPLETE* command button will initialize the form, increment *DOMAIN NUMBER* by one, and pass control to *DOMAIN VARIABLES*. At this point depending on the message box answers, the information is either appended to the list of domains in the table "Domain Key" or added to an empty table for domain one.

The *UNDO* command button will clear declared values and allow the domain to be re-defined prior to saving the definition (*DOMAIN COMPLETE*). The form is closed and the process ended by pressing *DONE*.

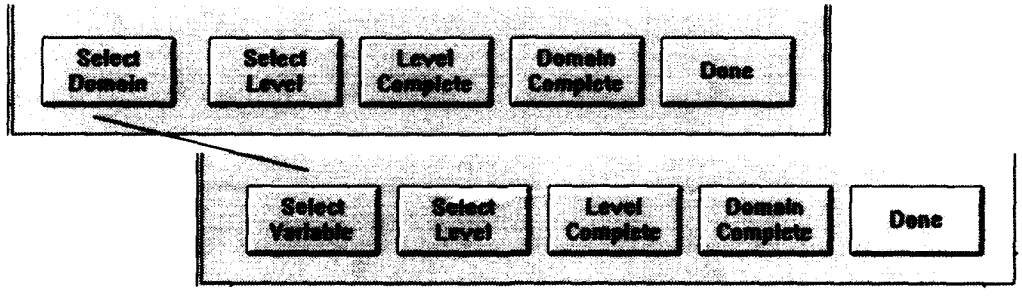
An example of a sampling domain definition is the following:

SERVICE	=	Navy
CONUS/OCONUS	=	OCONUS
RACE/ETHNICITY	=	non-Hispanic White
PAY GRADE	=	E1, E2, E3, E4

Re-Define Domains Form



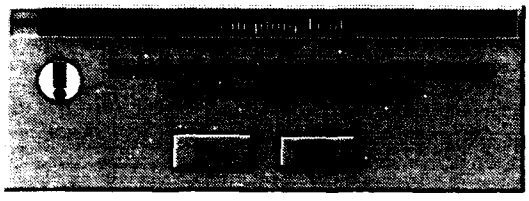
The form *Re-Define Domains* is used to alter the definition of one or more domains created previously. Type the relevant domain number into *DOMAIN NUMBER* and press *SELECT DOMAIN*. If the domain number is located in the list of reporting domains, then it's label is displayed in the *OLD DOMAIN LABEL* list box. The *SELECT DOMAIN* command button is replaced with the *SELECT VARIABLE* command button.



Select a domain variable with the mouse and press *SELECT VARIABLE*. The levels associated with this variable are listed in the *VARIABLE LEVELS* list box. Once a variable level is selected and *SELECT LEVEL* pressed, the label appears in the *NEW DOMAIN LABEL* list box and control returns to the level list.

Selection continues until all relevant values have been chosen (ex. E1-E4 for Pay Grade). *LEVEL COMPLETE* will return control to the *DOMAIN VARIABLES* list where the next variable associated with the domain may be chosen. Values for subsequent variables are listed in *VARIABLE LEVELS* and the process continues until the domain has been fully re-defined. Clicking the *DOMAIN COMPLETE* command button will eliminate the old domain definition from the table "Domain Key" and replace it with the updated version. The form will initialize and control is passed to *DOMAIN NUMBER*. The form is returned to the original state after the command button *SELECT VARIABLE* is replaced with *SELECT DOMAIN*.

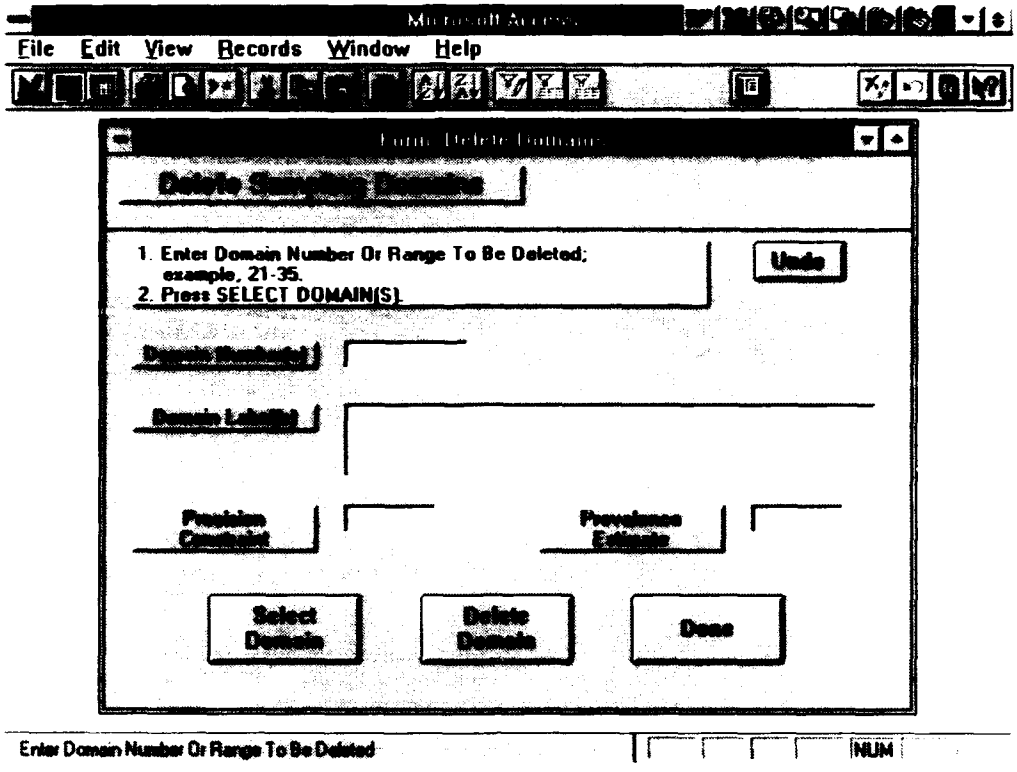
If the domain number is not found in the list of previously defined domains, then the number is either greater than the maximum defined domain number or has been deleted from the list.



If the number is greater than the maximum number in the list, then a two-button message box appears to determine if a new domain is to be created. If *YES*, then *DOMAIN NUMBER* initializes to the maximum domain number plus one and the *OLD DOMAIN LABEL* list box will contain "none" for the level. Declaration of this domain proceeds as described above. If *NO*, then the form resets and another domain number may be entered.

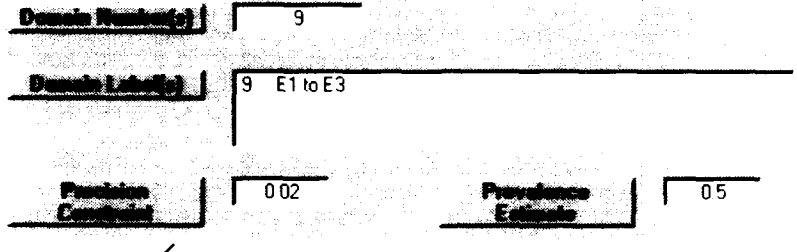
If the domain number is less than the maximum number in the list, then a message box appears indicating that the domain could not be found. The form will once again initialize. The *UNDO* command button will clear the current declared levels and allow the domain to be re-defined prior to saving the definition (*DOMAIN COMPLETE*). The form is closed and the process ended by pressing *DONE*.

Delete Domains Form



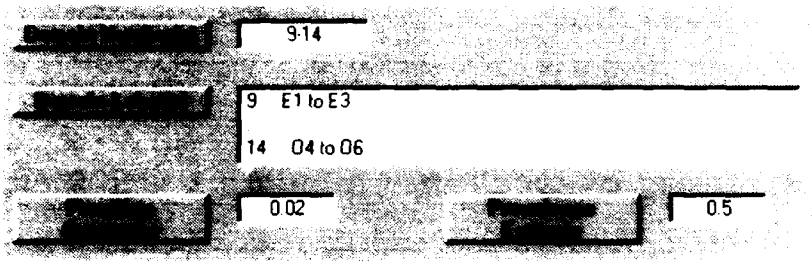
Another form used to modify the sampling domain list is the *Delete Domains Form*. Here, specified domain numbers are deleted from the table "Domain Key." Type the domain number or a single range of numbers into the *DOMAIN NUMBER(S)* text box and press *SELECT DOMAIN*.

If a single number is entered, the label is displayed in the *DOMAIN LABEL(S)* list box.



Values for the *PRECISION CONSTRAINT* and the *PREVALENCE ESTIMATE* are also displayed provided they have been defined.

If a range of domain numbers is entered into *DOMAIN NUMBER(S)*, then the first and last domain label is listed in *DOMAIN LABEL(S)*.



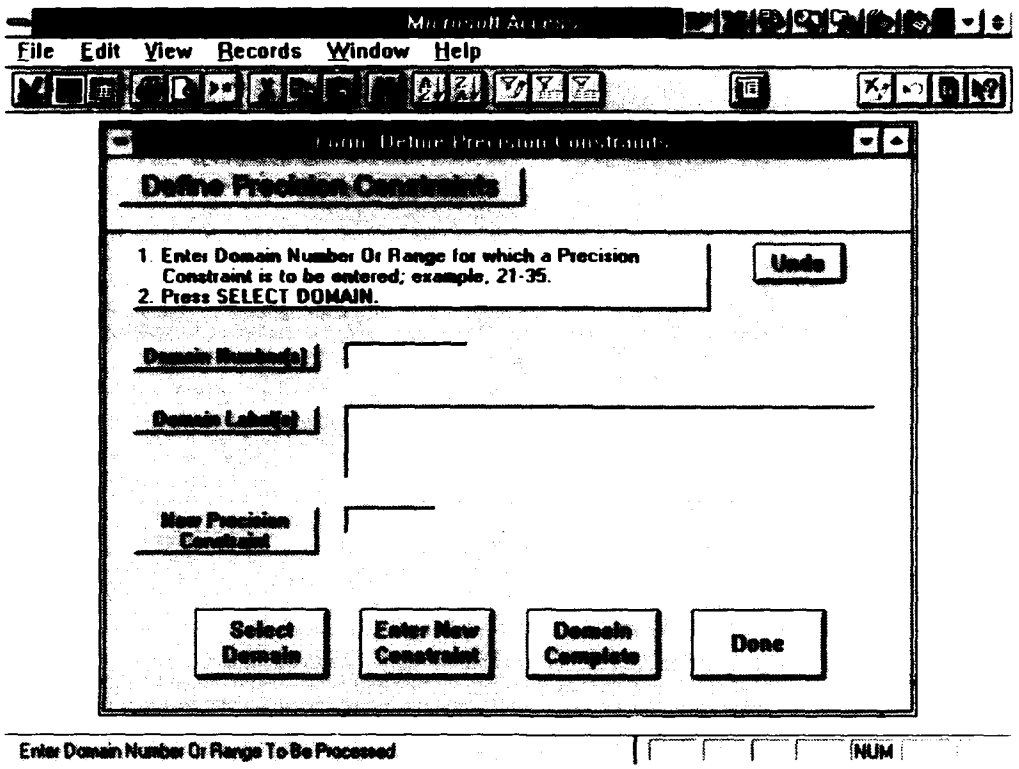
The *PRECISION CONSTRAINT* and *PREVALENCE ESTIMATE* are displayed for the first domain on the list.

Pressing *DELETE DOMAIN* will eliminate the domain or group of domains from the table. Control returns to *DOMAIN NUMBER(S)* for input of the next set of domain numbers to be processed. If none remain to be processed, press *DONE*. The *DONE* command button re-numbers the remaining domains in the table and closes the form.

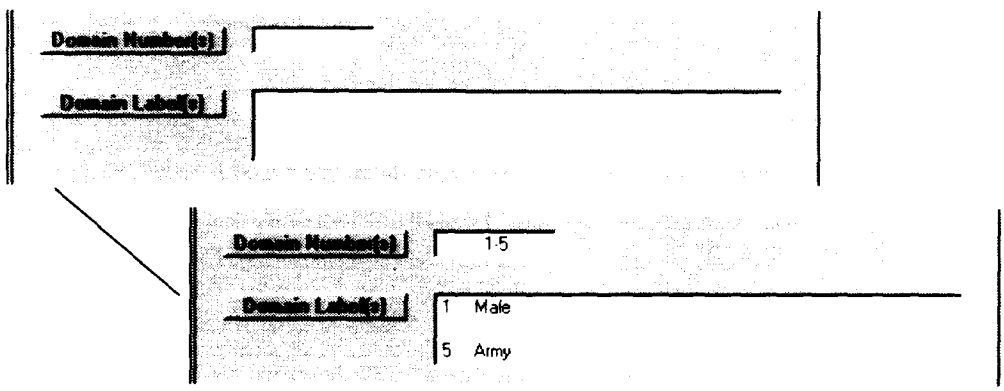
The *UNDO* command button will clear the current domain information from memory prior to pressing *DELETE DOMAIN*.

Remember that the domains have been re-numbered when examining future versions of the reports containing the domain information. The domain labels provided in the Tool reports aid in making the correct identification.

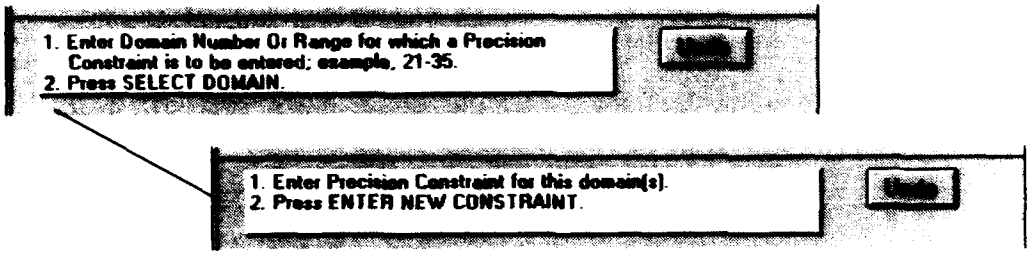
Define Precision Constraints Form



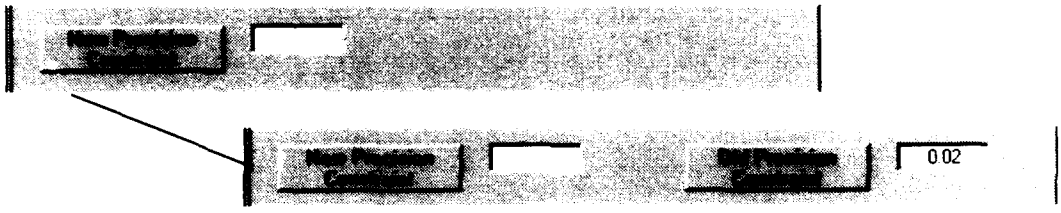
Precision constraints to be imposed on the reporting domains are entered using the *Define Precision Constraints Form*. Type a domain number or a single range of numbers into the *DOMAIN NUMBER(S)* list box and press *SELECT DOMAIN*. The label or labels associated with the first and last domain number are displayed in the *DOMAIN LABEL(S)* list box.



Simultaneously the instruction box changes to provide the instructions for the next step.



If the upper limit of the range is greater than the maximum domain number, then the process proceeds normally with the proper upper limit. If a constraint was defined previously for at least one of the domain numbers entered, then the form changes to additionally show the value of the previous constraint.



ENTER NEW CONSTRAINT is pressed after the new precision constraint is typed into the text box. This value should be greater than zero or equal to and less than or equal to one; a single-button message box appears if this criterion is not met.

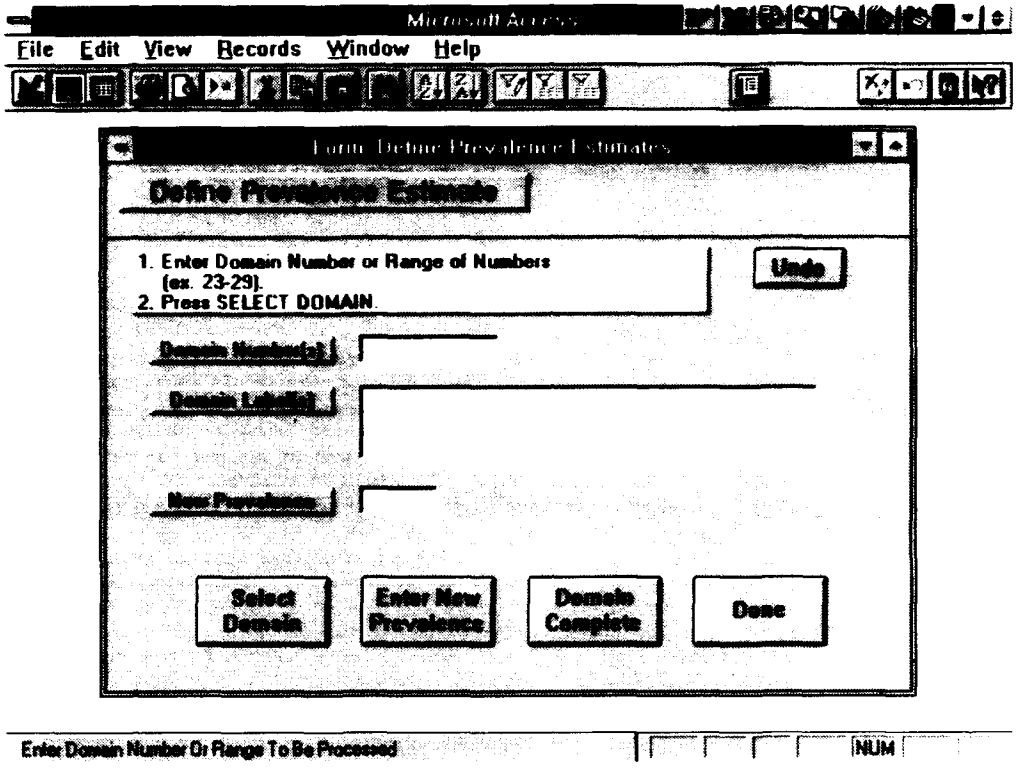
If the old precision information exists and this information is to be eliminated, enter zero for the *NEW PRECISION CONSTRAINT*.

A two-button message box will appear to verify that the old constraint should be deleted. If NO, the form resets and allows a new set of domain numbers to be entered into *DOMAIN NUMBER(S)*.

Either *DOMAIN COMPLETE* is pressed to save the constraint information to the table "Domain Key" or *UNDO* is pressed to restart the process.

DONE will close the form and return control to the calling menu.

Define Prevalence Estimates Form



Prevalence estimates for the reporting domains are entered using the *Define Prevalence Estimates Form* in the same way precision constraints are declared. Either a domain number or a single range of numbers is typed into *DOMAIN NUMBER(S)*. Pressing *SELECT DOMAIN* will cause the corresponding labels for the first and last domain number to be listed in the *DOMAIN LABEL(S)* list box.

If at least one prevalence estimate was defined previously for the domains, then this value will appear next to *NEW PREVALENCE* in the *OLD PREVALENCE* text box. The form at this point appears in much the same format as the *Define Precision Constraints Form* (pg. 37-38).

The updated estimate is entered into the *NEW PREVALENCE* text box and *ENTER NEW PREVALENCE* is clicked with the mouse. This value should be greater than or equal to zero and less than or equal to one; a single-button message box appears if this criterion is not met.

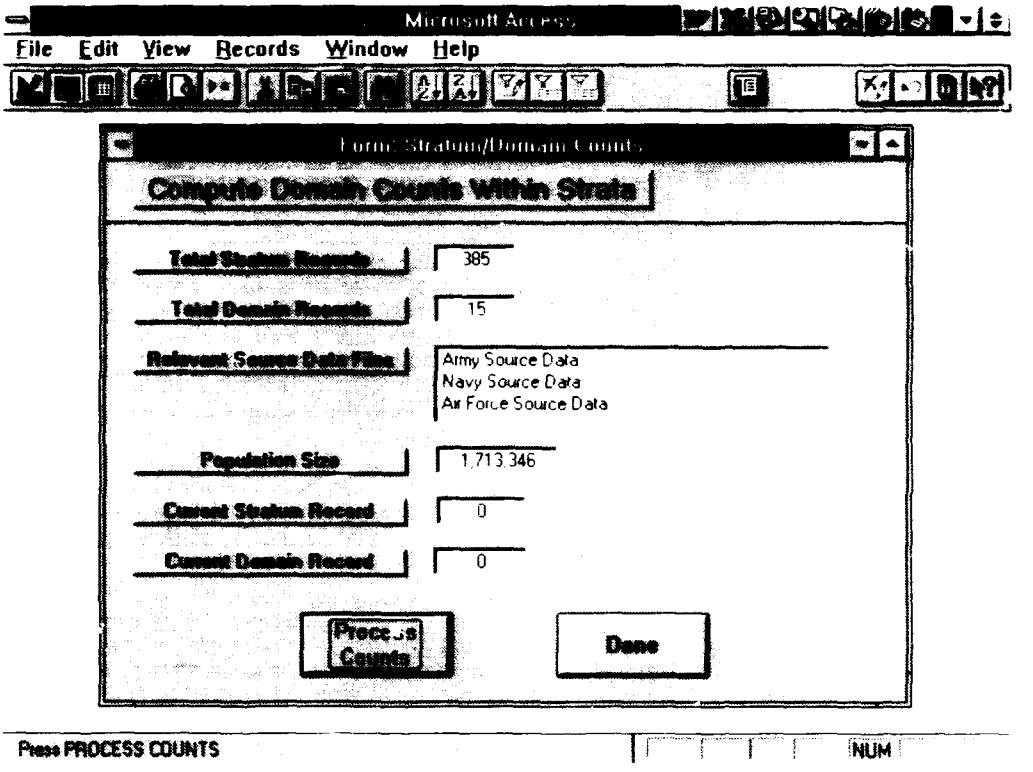
If an old prevalence estimate exists and this information is to be eliminated, enter zero for the *NEW PREVALENCE*. A two-button message box will appear to verify that the old data should be deleted and not simply re-defined. *UNDO* will clear the current information from memory and allow new prevalence estimates to be defined.

DOMAIN COMPLETE saves the information to the "Domain Key" table and resets the form for the next set of domains.

DONE closes the form and returns control to the *Reporting Domains Menu*.

For an example, previous studies have estimated that 55% of females in the armed forces have reported experiencing at least one occurrence of some specified event. Domains 23-28 all possess Gender=Female in their definitions. Thus 0.55 is entered in *NEW PREVALENCE* for *DOMAIN NUMBER(S)* 23-28.

Stratum/Domain Counts Form

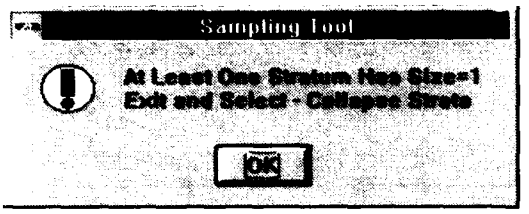


The *Stratum/Domain Counts Form* is used to compute domain sizes within strata and marginal domain sizes on the frame data file or files. A discussion of the total time needed to compute the stratum/domain sizes is thought to be necessary before explaining the form.

Once processing has started the system will be in use for an extended period of time. Total processing time differs with several factors: type of system, the size and the number of source data files, the number of strata, the number of domains, and number of stratum/domain counts to be computed. Approximately, 4 hours were used to process stratum/domain counts for a single source data file with n=8,824 records, n=389 strata, and n=200 domains on a 486 computer operating at 33 MHz. This time was approximately doubled on a 386 computer operating at the same speed, and somewhat less than halved on a Pentium processor operating at 90 MHz. Therefore, it is best to run the process when the system is not needed for another task.

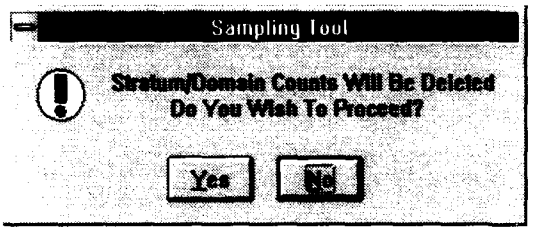
One method used to trim the processing time is to only compute those counts for domains which are newly created or re-defined. Thus if 150 domains counts were calculated in the first sampling tool session and 20 domains were added to the list in a subsequent session, only 389 x 20 (=7,780) stratum/domains counts would be computed instead of 389 x 170 (=66,130) domains. The procedure will assist in but will not alleviate the processing time problem. This should be kept in mind prior to invocation of the form.

When the form is first invoked, the sampling tool scans through the current stratum size table to determine if any strata have a size of one.



If at least one is found, then a single-button message box will appear to indicate the problem. Pressing OK will return control to the *Reporting Domains Menu*. Examine the *Stratum Sizes Report* to determine the strata that need to be collapsed and then invoke the *Collapse Strata Form*.

After examining the stratum sizes, the sampling tool determines if a table of stratum by domain counts was created previously.



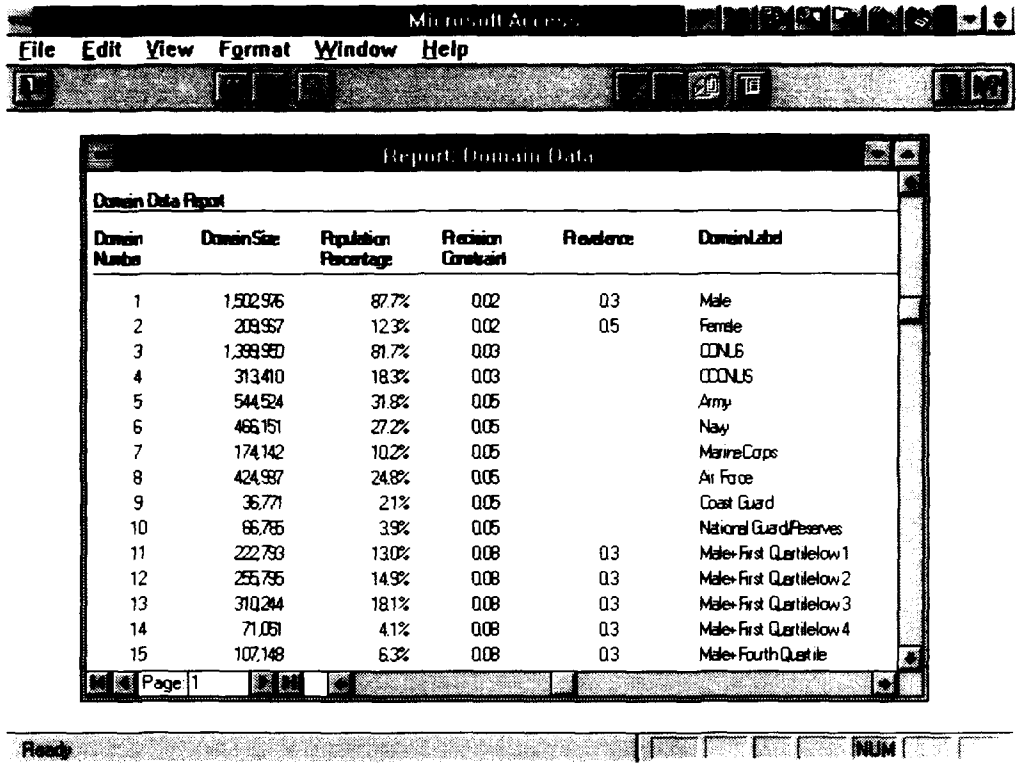
If so, a two-button message box appears to determine whether or not this table should be deleted and re-created.

An option other than deleting the old information would be to select *NO*, exit the sampling tool to the table definitions list, and copy the table "Stratum/Domain Counts" to a new name. Therefore, only a second version of the table is deleted. Motivation for saving the earlier version of the table is the computer time needed to re-create it.

Once the form is activated, the total number of strata constructed in the *Stratification Variables* and the *Stratum Levels* forms is listed in the *TOTAL STRATUM RECORDS* text box. The number of records appearing in the "Domain Key" table is displayed in the *TOTAL DOMAIN RECORDS* text box. The data sets listed for the *RELEVANT SOURCE DATA FILES* are all of the files processed in the form *Construct Strata* in the calculation of the sampling strata counts. These pieces of information are useful in estimating the time needed to accomplish the step. The text box *POPULATION SIZE* is the total frame count.

When the form is invoked, press *PROCESS COUNTS*; this will start the procedures for calculating the relevant counts. Note, this form does not contain an *UNDO* command button. A blue meter appears in the lower left corner to show the percent of records out of the total that have been processed. The *CURRENT STRATUM RECORD* and the *CURRENT DOMAIN RECORD* are incremented while each file is being processed. When all of the operations have been completed, the system will beep and control goes to the *DONE* command button. If records have been processed, *DONE* will save the information to the table "Stratum/Domain Counts," close the form and return control to the *Reporting Domains Menu*. If this command button is pressed prior to processing (*PROCESS COUNTS*), a two-button message box will appear to determine if the user wishes to close or reset the form.

Domain Data Report



Domain Number	Domain Size	Population Percentage	Precision Constraint	Prevalence	Domain Label
1	1,502,976	87.7%	0.02	0.3	Male
2	209,957	12.3%	0.02	0.5	Female
3	1,399,950	81.7%	0.03		CONUS
4	313,410	18.3%	0.03		OCONUS
5	544,524	31.8%	0.05		Army
6	466,151	27.2%	0.05		Navy
7	174,142	10.2%	0.05		MarineCorps
8	424,937	24.8%	0.05		Air Force
9	36,771	2.1%	0.05		Coast Guard
10	66,785	3.9%	0.05		National Guard/Reserves
11	222,793	13.0%	0.08	0.3	Male+First Quartilelow 1
12	255,795	14.9%	0.08	0.3	Male+First Quartilelow 2
13	310,244	18.1%	0.08	0.3	Male+First Quartilelow 3
14	71,051	4.1%	0.08	0.3	Male+First Quartilelow 4
15	107,148	6.3%	0.08	0.3	Male+Fouth Quartile

Once at least one sampling domain has been defined, the *Domain Data Report* may be invoked. The reporting domains are identified by *DOMAIN NUMBER* and *DOMAIN LABEL*. If the domain counts or sizes have been computed (*Stratum/Domain Counts Form*), then those values are provided in the *DOMAIN SIZE* column. Otherwise, the column remains blank. *POPULATION PERCENTAGE*s (reporting domain size / total population size) are calculated only if the domain counts exist.

PRECISION CONSTRAINTS and *PREVALENCE ESTIMATES* are listed for each domain. Note that prevalence estimates are missing for domains 3-10 since these values have not been defined. The left-most button on the formatting toolbar will close the report and return control to the calling menu. Alternately clicking the printer icon in the Access toolbar prints the report.

Building The Cost Model

A cost model, in general, describes the total cost of a survey in terms of two components, fixed costs and variable costs. Fixed costs are those that are unaffected by the size of the sample used in the survey. Examples of fixed costs might include progress and financial reporting. Conversely, variable costs are those that do depend on the sample size selected, such as data collection activities. Fixed costs do not enter into the sample allocation calculations. Hence, unless another reason exists for including fixed costs, the cost modeling exercise can be restricted to the variable cost component.

A cost model is developed by first compiling an exhaustive list of all of the activities to be undertaken to carry out the survey. Then cost estimates are developed for each item on the list, perhaps drawing on recent experience with more or less similar surveys. The items comprising the list can, for convenience in developing the list and ensuring that it is exhaustive, be classified into groups of related activities. Thus the variable cost of any survey can potentially involve activities associated with;

- constructing the sampling frame,
- selecting the sample,
- developing the survey instruments,
- collecting the data,
- editing the data,
- processing the data,
- analyzing and reporting the data.

Some arbitrariness exists, of course, in associating a given activity with one of these groups. Some investigators might classify the costs of reproducing the survey instruments, for example, as a data collection cost rather than an instrument development cost. Others might prefer the reverse. Further, some activities might include both fixed and variable costs, necessitating some apportioning of the activity into the two groups.

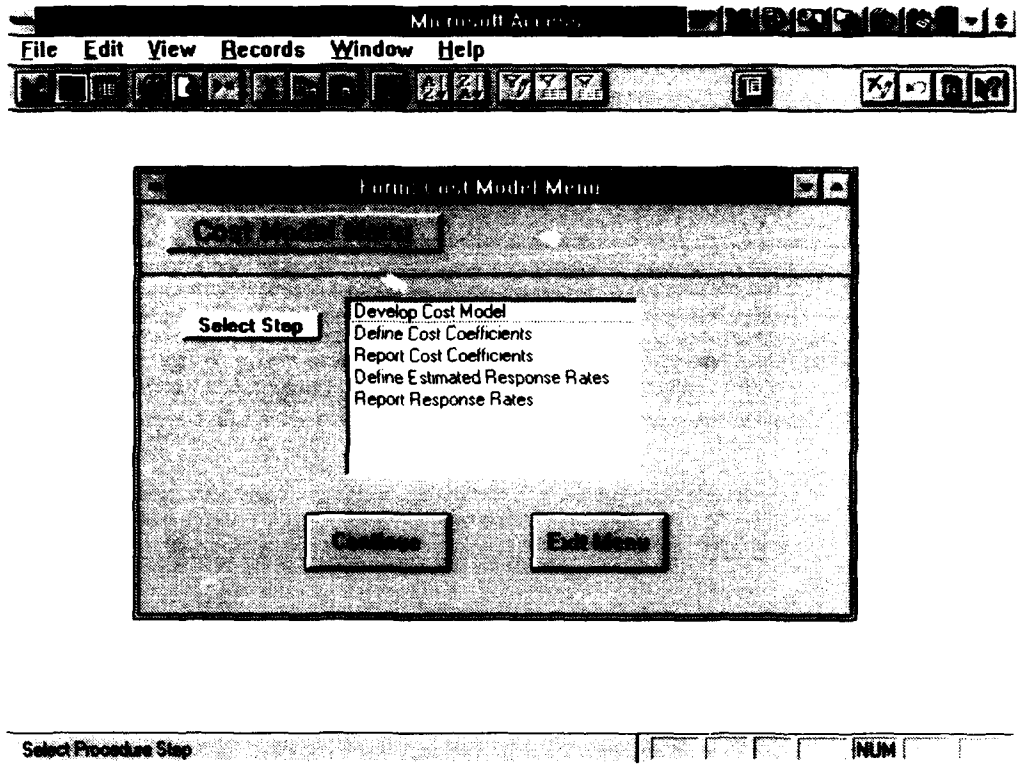
The cost model capability of this version of the Tool is restricted to mail data collection procedures applicable to a stratified random sampling design. Given this restriction, costs associated with frame construction, sample selection, developing survey instruments, and analyzing the data are reasonably considered to be fixed costs, not dependent on the size of the sample selected. Accordingly the default groups of variable cost activities assumed by this version of the Tool consist of collecting, editing, and processing the data. These defaults are easily changed should they be inappropriate.

The objective of the modeling exercise becomes one of quantitating the per observation average stratum-level costs and supplying this information to the Tool. In this context an observation is defined as a sample person for whom the information needed to compute a parameter estimate has been obtained. In context, an observation is obtained if a sample person returns a usable questionnaire. However, eligibility issues may also be involved in a given survey. An ineligible person by definition is a sample person who is not a member of the population of inferential interest. A sample person who is known to be ineligible contributes an observation and the costs of determining eligibility status, perhaps incurred in an initial screening survey, are included in the per observation stratum-level averages.

Data collection costs for a mail survey are commonly computed using the per unit cost of a mailing rather than a per observation cost. Input to the Tool therefore is of two types, the per observation costs of such activities as data processing and editing, and the per package cost of each mailing. One mailing is distinguished from another if the experienced response patterns are used to determine the number of packages mailed on different occasions. For example, a lead letter followed by a questionnaire followed by a thank-you-reminder postcard sent on different occasions to the same set of persons counts as one mailing. Conversely an additional questionnaire sent only to persons tabulated as non-respondents at some point in the data collection period counts as an additional mailing. The difference is that subsequent mailings are made only to non-respondents to a previous mailing.

The Tool computes per observation data collection costs given per unit mailing costs for an arbitrary number of mailings. That is, after entering the number of mailings and the expected response rates to each mailing, the user is queried for the per unit cost of each mailing and the Tool computes the required per observation cost.

Cost Model Menu

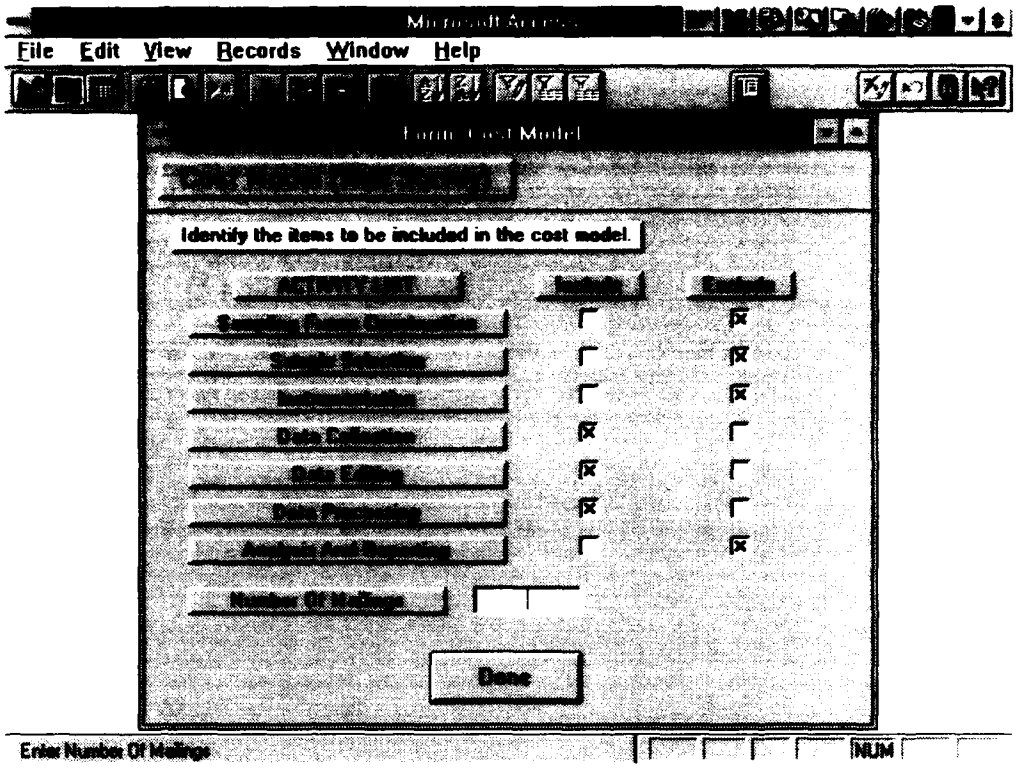


The cost model procedures are accessed from the *Cost Model Menu* illustrated above. As with other menus, a form or report is invoked by clicking the corresponding menu item and pressing *CONTINUE*.

Note that this menu contains items involving both cost coefficients and response rates. Both types of information are needed to develop the cost model.

The *EXIT MENU* command button is pressed to close the menu and return control to the *Sampling Tool Menu*.

Cost Model Form



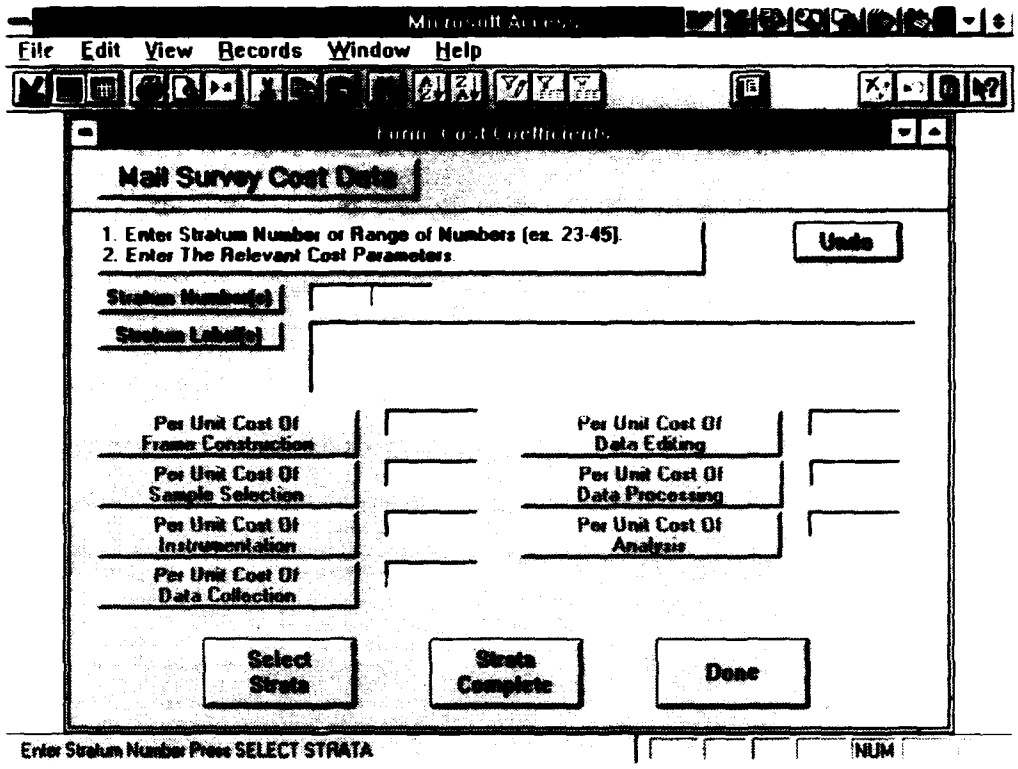
The cost components to be included in the cost model are defined in the *Cost Model Form*. When the form opens, the default components for a mail survey cost model are indicated under the *INCLUDE* column on the form. The defaults may be accepted or modified by clicking the appropriate box under either the *INCLUDE* or *EXCLUDE* column.

The components included should encompass the variable or non-fixed cost portion of the survey. That is, the cost coefficients defined in a subsequent form (*Cost Coefficients Form*) should depend entirely on the number of units being handled within that activity. Thus the coefficient associated with the *DATA EDITING* activity, for example, should be provided on a per completed questionnaire basis.

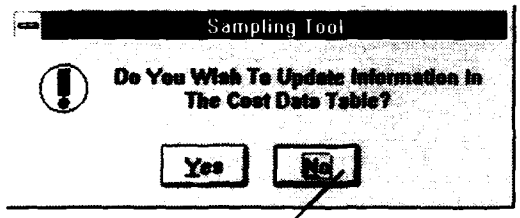
The per unit *DATA COLLECTION* costs are computed by the Tool using the user-supplied cost of each mailing. The number of mailings is entered in the *NUMBER OF MAILINGS* text box on this form. A single-button message box will appear indicating that this value is required if the *DONE* command button is pressed prematurely.

DONE causes the *ACTIVITY* LIST information to be saved to the table "Model Items" and closes the form. The activities to be included in the cost model may be changed at any time prior to pressing *DONE*.

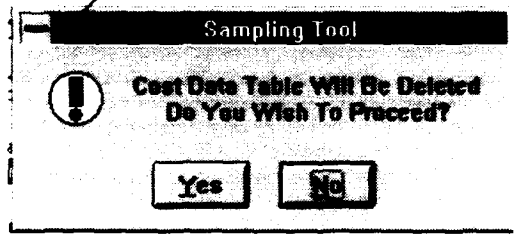
Cost Coefficients Form



The stratum-specific coefficients of the cost model components defined in the *Cost Model Form* are entered using the *Cost Coefficients* form. Before the form is opened the Tool first determines if the "Cost Data" table exists. If the table does not yet exist, then the form appears as shown above.



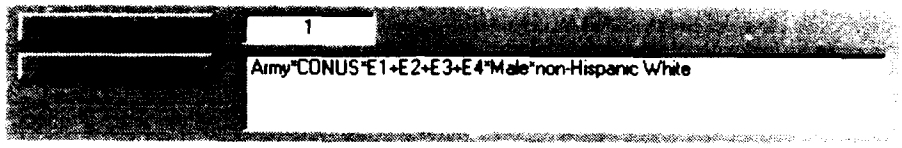
However, if any cost information was specified previously, the form does not open immediately. A two-button message box appears inquiring if the existing cost data are to be updated (*YES*) or completely re-defined (*NO*).



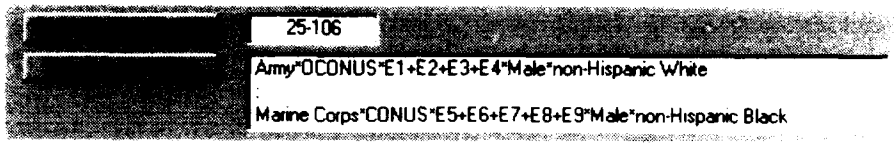
A second two-button box will appear if the data are to be re-defined to verify that the "Cost Data" table may be deleted. *NO* exits the form with no further action.

Regardless of the circumstances, the form opens with the *STRATUM NUMBER(S)* text box highlighted. Enter a stratum number or a single range of stratum numbers as appropriate and press the *SELECT STRATA* command button.

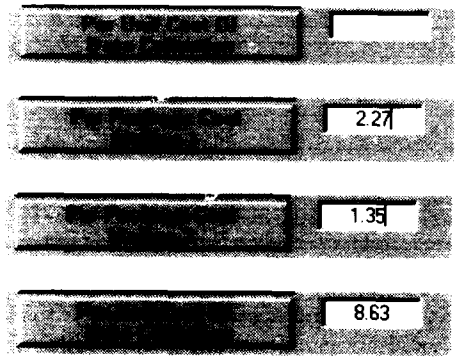
If a single stratum number has been selected the Tool responds by listing the label for the stratum in the *STRATUM LABEL* list box.



If a range of numbers has been selected, the labels for the first and last stratum in the range are shown.



Control is passed to the PER UNIT text box corresponding to the first listed activity included in the cost model. Cost items identified as excluded from the cost model appear greyed out on the form and cannot be activated. The cost coefficients may be entered as either a number or as currency using a dollar sign. The values typed in the text box are "entered" either using the enter (return) key on the keyboard or by clicking another of the cost items included in the model.



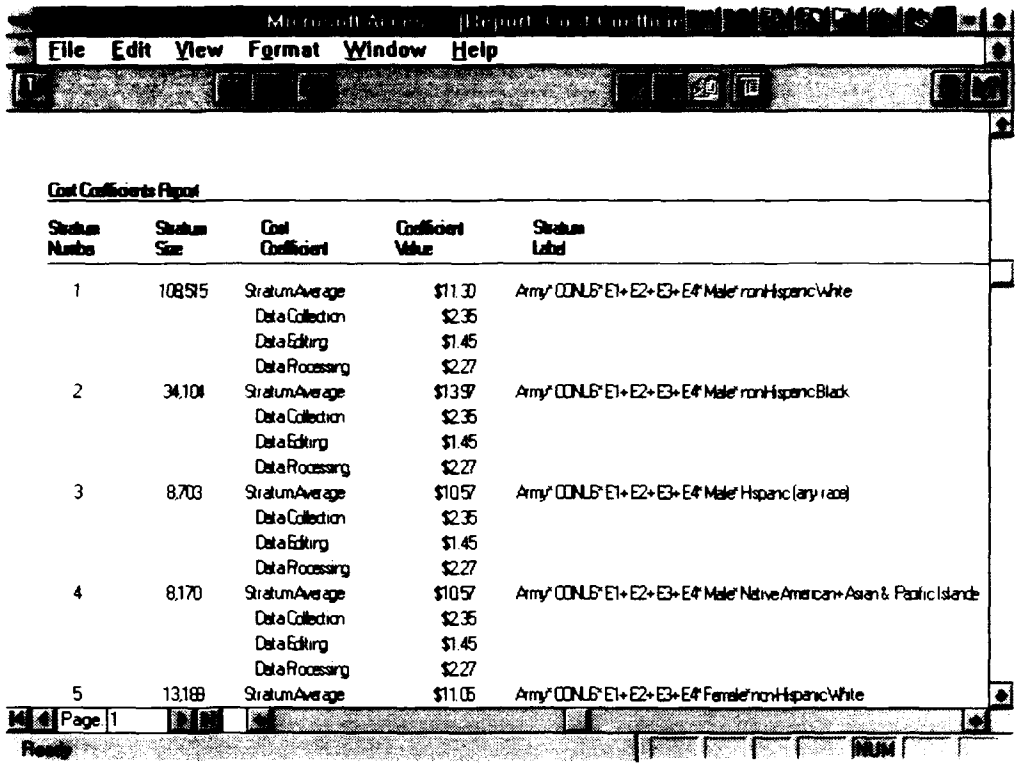
When the *PER UNIT COST OF DATA COLLECTION* text box is first activated its label changes to *PER PACKAGE COST MAILING 1*. Enter the mailing 1 cost and press enter on the keyboard. The label changes to *MAILING 2*, and so on, through the number of mailings entered on the *Cost Model* form. When cost coefficients have been entered for all of the mailings, the label reverts to *PER UNIT COST OF DATA COLLECTION* and the text box displays the per unit cost computed using the per package mailing costs.

The above sequence occurs, however, only if the response rate information required to compute the *PER UNIT COST* coefficient has been previously entered (see Response Rates Form, pgs. 53-55). This would normally be the case if the cost coefficients are being updated but would not necessarily be the case if the coefficients are being entered for the first time. In the response rate information is missing, then the last per-mailing cost value and label remain. Control passes to the next component in the cost model.

When all of the relevant per unit cost coefficients have been entered, pressing the *STRATA COMPLETE* command button saves the information and resets the control to the *STRATUM*

NUMBER(S) text box, anticipating further modifications. *UNDO* clears all entries in the form and returns control to the *STRATUM NUMBER* text box. Pressing *DONE* closes the form and returns control to the calling menu.

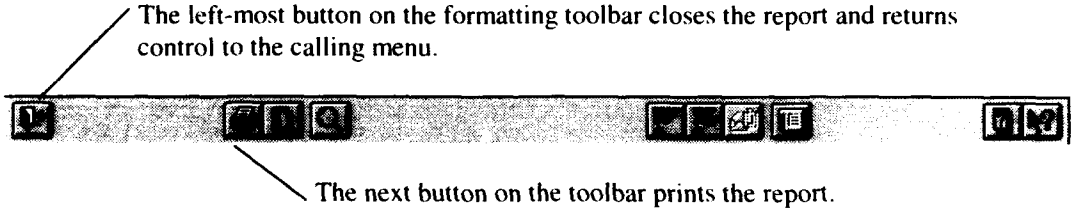
Cost Coefficients Report



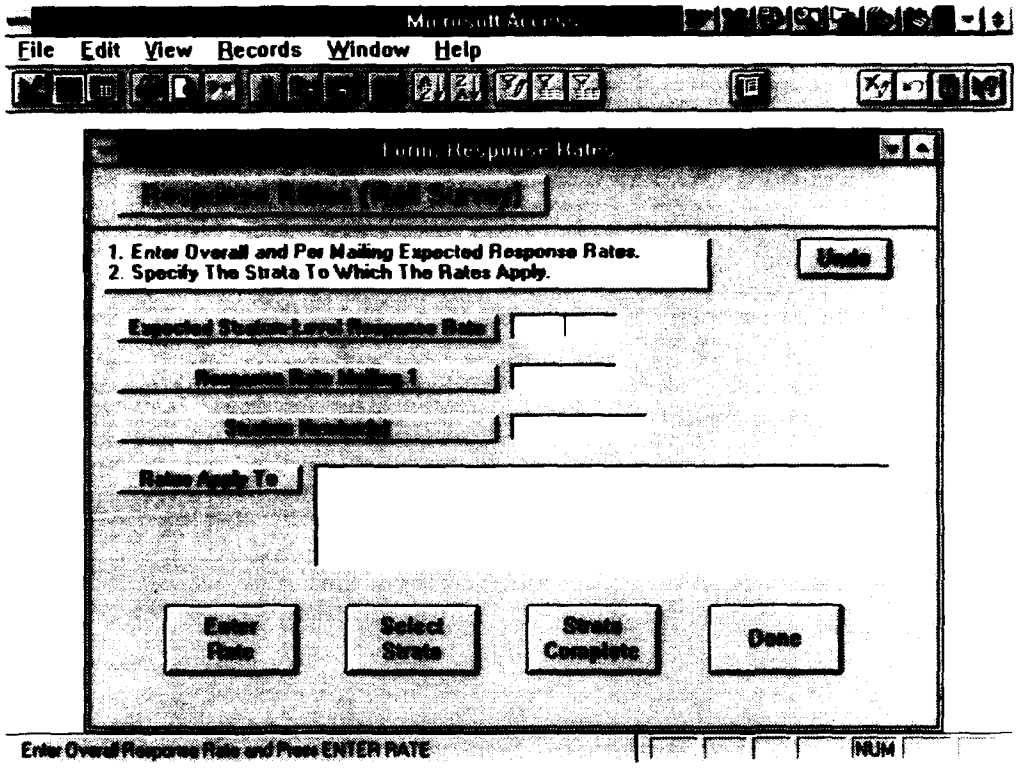
Stratum Numbe	Stratum Size	Cost Coefficient	Coefficient Value	Stratum Label
1	108,515	StratumAverage	$11.30	Army* CONUS* E1+E2+E3+E4* Male* nonHispanic White
		Data Collection	$2.35	
		Data Editing	$1.45	
		Data Processing	$2.27	
2	34,101	StratumAverage	$13.97	Army* CONUS* E1+E2+E3+E4* Male* nonHispanic Black
		Data Collection	$2.35	
		Data Editing	$1.45	
		Data Processing	$2.27	
3	8,703	StratumAverage	$10.57	Army* CONUS* E1+E2+E3+E4* Male* Hispanic (any race)
		Data Collection	$2.35	
		Data Editing	$1.45	
		Data Processing	$2.27	
4	8,170	StratumAverage	$10.57	Army* CONUS* E1+E2+E3+E4* Male* Native American+Asian & Pacific Islande
		Data Collection	$2.35	
		Data Editing	$1.45	
		Data Processing	$2.27	
5	13,188	StratumAverage	$11.05	Army* CONUS* E1+E2+E3+E4* Female* nonHispanic White

The per unit (observation) average stratum costs are summarized in the *Cost Coefficients Report*. The components "included" in the cost model using the *Cost Model Form* are listed in the *COST COEFFICIENT* column. Note, the *STRATUM AVERAGE* value can only be defined after all cost coefficient and response rate information has been entered into the "Cost Data" table.

Coefficients for the cost model components are listed on the report under the *COEFFICIENT VALUE* column. The *STRATUM SIZE* and the *STRATUM LABEL* are listed to aid in making comparisons across several reports.



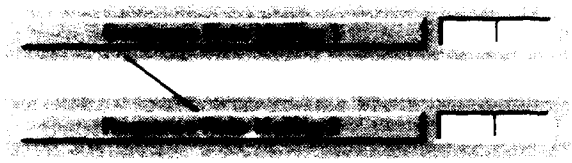
Response Rates Form



The *Response Rates Form* is used to define the average stratum-level response rates and the response rate to each mailing. The form opens with the *EXPECTED STRATUM-LEVEL RESPONSE RATE* text box highlighted. Key the response rate expected to be obtained over all of the mailings and press the *ENTER RATE* command button.

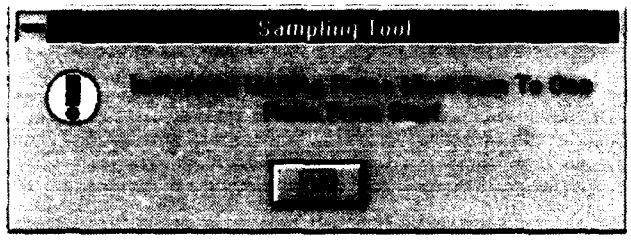
The values keyed must be in decimal form within the range of 0 to 1. If the value does not meet these conditions a single-button message box will invoke. Click *OK* to return to the form to edit the value.

After the overall response rate has been entered, control passes to the *RESPONSE RATE MAILING 1* text box. Key the proportion of the stratum-level response rate that is expected to be obtained on the first mailing and click the *ENTER RATE* command button.



If the survey involves more than one mailing, the text box label is updated to receive the response rate for the next mailing.

Note that because the response rate keyed for each mailing is the proportion of the stratum-level response rate, the sum of the rates over all mailings must equal one.



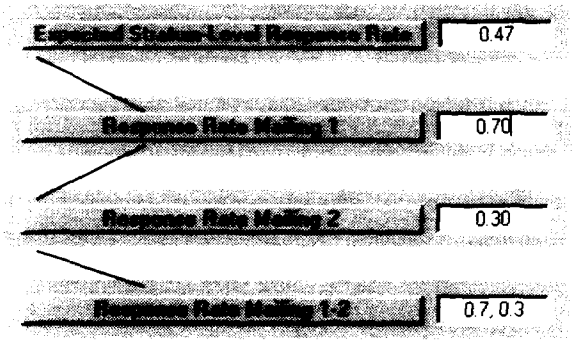
If the mailing rates do not sum to one, a single-button message box appears. Click *OK* to return to the form. Re-key the rates starting with mailing 1. The partition is automatic if the study has only one mailing.

When the proportion for the last mailing is entered, the text box label shows the number of mailings and the rates themselves are listed in the text box separated by commas.

Control is then passed to the *STRATUM NUMBER(S)* text box. Enter the stratum number or a single range of stratum numbers to which the response rates just entered are to be applied. Provided the stratum information is found in a table, labels for the first and last stratum are listed in the *RATES APPLY TO* list box. *STRATA COMPLETE* will save the stratum-level information and reset the form This process continues until the *DONE* command button is clicked. *DONE* closes the form once all of the information has been saved.

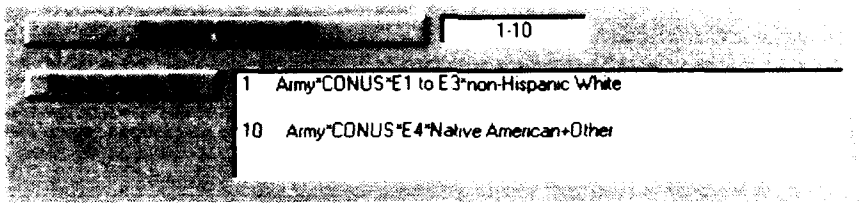
The *UNDO* command button will erase the data from the screen and initialize the form only if it is pressed before the *STRATA COMPLETE* button.

Suppose by way of example that the response rates for strata 1-10 are expected to be 47% and two mailings will be sent to the sampled individuals. An estimated 70% of the completed questionnaires will be received from the first mailing.



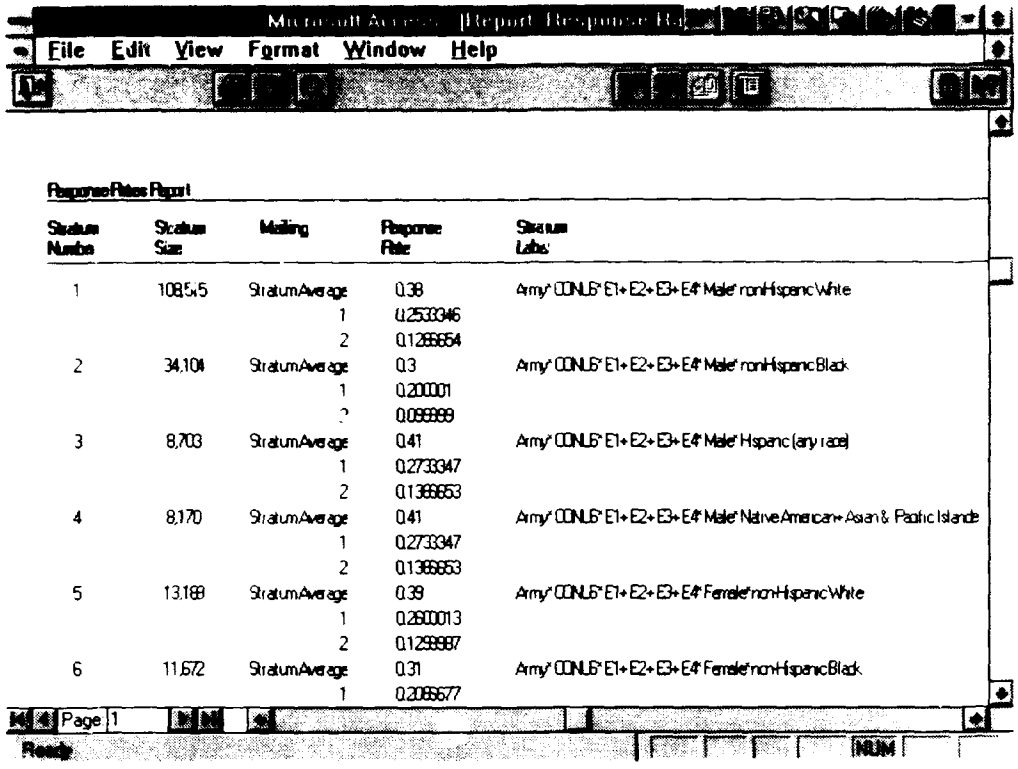
Type 0.47 into the *EXPECTED STRATUM-LEVEL RESPONSE RATE* text box and press *ENTER RATE*. Then type 0.70 and 0.30 in turn into the *RESPONSE RATE MAILING 1* and *MAILING 2* text boxes, pressing *ENTER RATE* after each entry.

Control is passed to *STRATUM NUMBER(S)*. The string "1-10" is entered for the stratum range and *SELECT STRATA* is clicked. The stratum labels for stratum 1 and stratum 10 are displayed.



Press the *STRATA COMPLETE* command button to save the information, reset the form and pass control back to the *EXPECTED STRATUM-LEVEL RESPONSE RATE* text box to process another set of strata.

Response Rates Report



The response rates by stratum are summarized in the *Response Rates Report*. The values for the *RESPONSE RATEs* are provided for each *MAILING* within *STRATUM NUMBER*. The *STRATUM AVERAGE* is the sum of the per-mailing response rates.

Recall that the response rates are used to compute the stratum-level cost coefficient (see *Cost Coefficients Report* pg. 52). Once the *STRATUM AVERAGE* response rate has been calculated, then the *STRATUM AVERAGE* cost coefficient on the *Cost Coefficients Report* is calculated. The *STRATUM LABEL* has been provided for ease in comparison across several reports.

The left-most button on the formatting toolbar closes the report and returns control to the calling menu. The next button on the toolbar prints the report.

Computing The Sample Allocation

Using the information supplied to this point the Tool computes the least cost allocation of the sample that will simultaneously satisfy the imposed precision requirements. A distinction is made between the number of observations required to satisfy the precision requirements and the sample size needed to provide the required number of observations. The Tool first computes the number of observations needed in each stratum and then inflates these using the stratum-level response rate information to obtain the required stratum-level sample sizes. Both values are reported.

An iterative numerical algorithm is used to obtain the allocation solutions. A general discussion of this iterative method is given in a paper by Dr. James R. Chromy (1987). An applied discussion using specific examples from the SAFS is given in a paper by Dr. Robert E. Mason and others (1995). The user is required to input the decision to be used to terminate the allocation procedure. The termination decision can be stated in one of two ways. One way is to simply specify the maximum number of iterations the procedure is to use. A second way is based on the Karush-Kuhn-Tucker necessary conditions for a solution to be obtained to a set of constrained non-linear equations. The Tool computes a numeric quantity, referred to as the convergence criterion, that approaches zero over an indefinite number of iterations. Rather than specifying the number of iterations, the user may alternately specify how closely the convergence criterion must approach zero before the procedure is terminated.

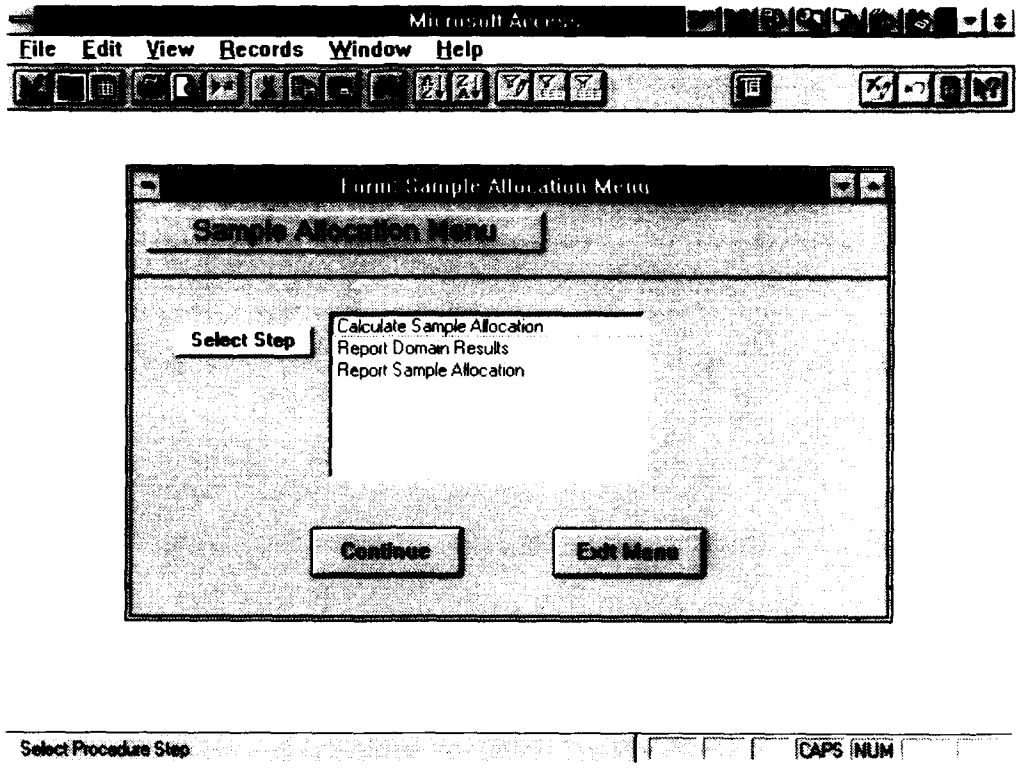
As well as reporting the finally determined sample allocation solutions, the Tool provides summary information that is useful in modifying the precision constraints. A common experience in designing a sample is to find that the initial precision constraints need to be modified to bring the total sample size in line with budget realities. Less commonly the initial constraints may require some tightening. The *Domain Results Report* lists two quantities that are useful in this regard.

First, in setting up the numerical algorithm the Tool computes generalized Lagrange multipliers that will satisfy each of the precision constraints individually (rather than jointly). A comparison of these initial values with the final Lagrange multiplier values identifies those constraints that are the most important in determining the allocation solutions, and by implication, the variable survey cost. The Tool identifies the ten most important constraints on a scale of 1 to 100. Even a small relaxation of one or more of the identified constraints can produce a sizable reduction in cost.

Second, the Tool computes the total design effect for each of the reporting domain estimates. The design effect is simply the ratio of the sampling variance given the actual sampling design divided by the variance that would be obtained using a simple random sample with the same number of observations.

Reporting domains with the largest design effects will be those that include subdomains that have been oversampled. For example, if female company grade officers are oversampled relative to other domains, then any more inclusive domain that contains females or company grade officers will have a relatively large design effect. Excessively large design effects might lead an investigator to modify the stratification scheme, modify the domain definitions, or relax the precision constraints for some of the subdomains.

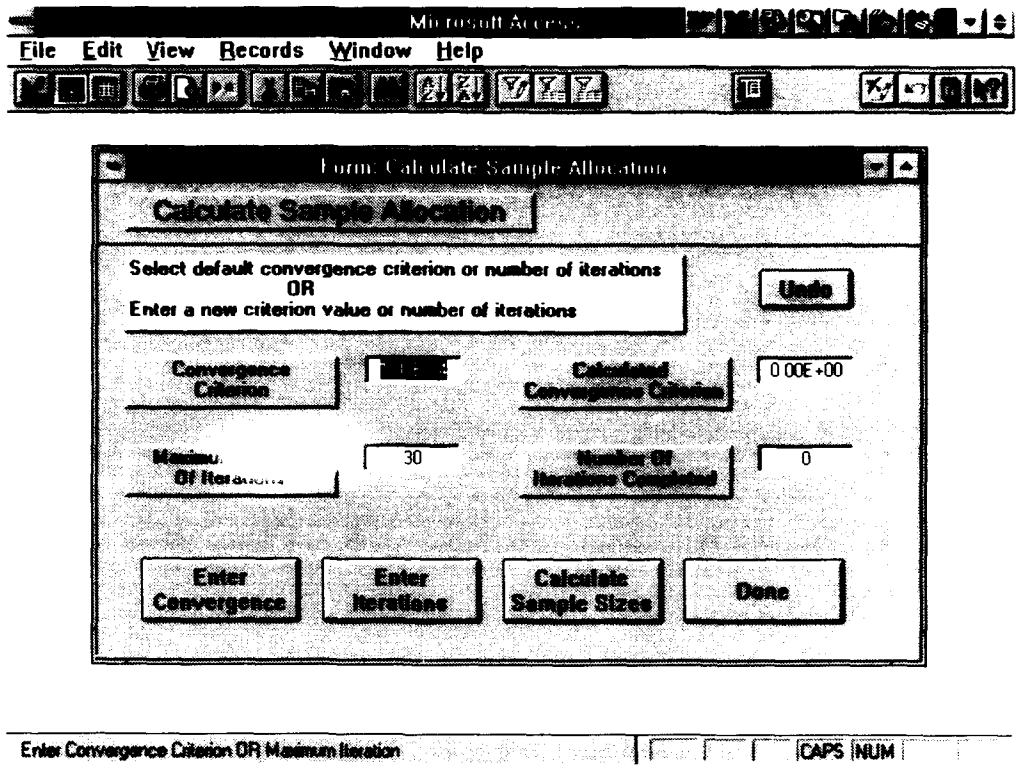
Sample Allocation Menu



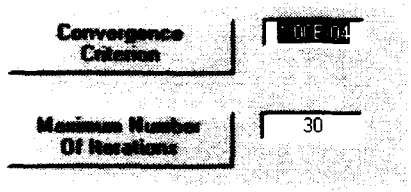
The final task within the Sampling Tool is the calculation of the sample allocation. This procedure is completed in the *Calculate Sample Allocation Form* called by the *Sample Allocation Menu*. The form and two reports are invoked by clicking a line item with a mouse and pressing *CONTINUE*.

The *EXIT MENU* command button is pressed to close the menu and return control to the originating menu, the *Sampling Tool Menu*.

Calculate Sample Allocation Form



Sample allocations are calculated using the *Calculate Sample Allocation Form*. An iterative numerical procedure is used to calculate the allocation solutions. The procedure is terminated by specifying a value in either the *CONVERGENCE CRITERION* text box or the *MAXIMUM NUMBER OF ITERATIONS* text box.

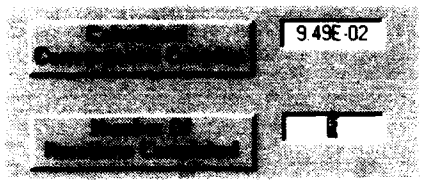


The form opens with a default *CONVERGENCE CRITERION* of 0.0001. More accurate allocation solutions are obtained if a value closer to zero is entered. Alternately the default number of iterations can be increased to improve the accuracy of the solutions.

The default values will usually provide three digit accuracy in the results. To accept the default *CONVERGENCE CRITERION* press the *ENTER CONVERGENCE* command button. To accept the default *MAXIMUM NUMBER OF ITERATIONS* click the text box and then the *ENTER ITERATIONS* command button.

To enter new values in either text box, click the box, enter the new value and then press the appropriate command button. Pressing *UNDO* will initialize the form with the default values.

Press *CALCULATE SAMPLE SIZES* once either a *CONVERGENCE CRITERION* or a *MAXIMUM NUMBER OF ITERATIONS* has been specified.



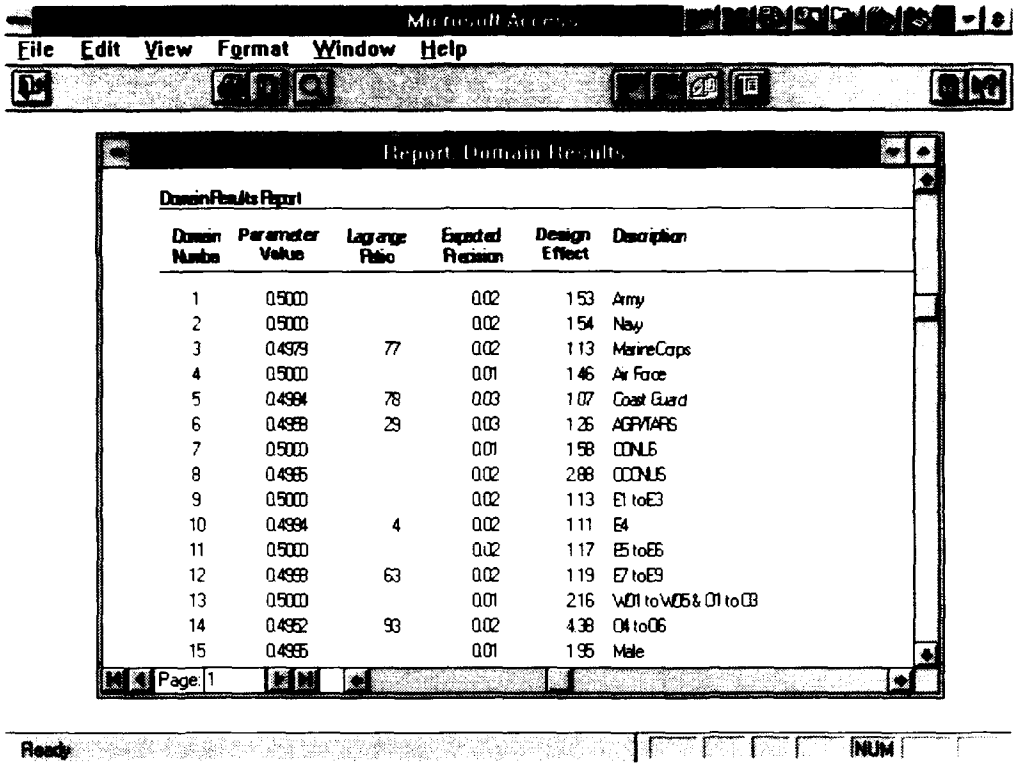
As the process iterates, the *CALCULATED CONVERGENCE CRITERION* and *NUMBER OF ITERATIONS COMPLETED* text boxes are updated.

The steps in the numerical algorithm are described in the status line with a meter that shows the relative progress of each step.

Relevant information is saved to the tables "Domain Information" and "Sample Allocation."

Control passes to the command button *DONE* upon completion of the procedure. *DONE* closes the form and returns control to the *Sample Allocation Menu*.

Domain Results Report



The first report to display values calculated from *Calculate Sample Allocation Form* is the *Domain Results Report*. Domain information is listed by *DOMAIN NUMBER* and the domain label, *DESCRIPTION*.

The domain estimate is shown in the *PARAMETER VALUE* column. The value is computed by applying the user-supplied prevalence to the actual domain sizes in each stratum and then re-computing the domain level prevalence based on the actual size of the domain. Hence the values shown in the table may not agree exactly with prevalences specified using the *Define Prevalence Estimates* form.

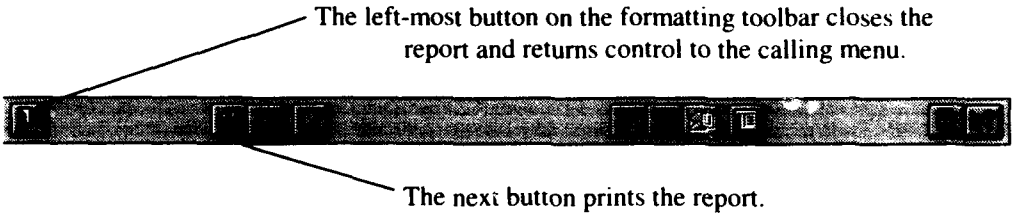
The *LAGRANGE RATIO*s are used to identify those precision constraints which have the greatest effect in determining the survey costs. Values are shown on a scale from 1 to 100 for the top ten most important precision constraints. A value of 100 implies that no reduction in the number of observations needed to satisfy the requirement has accrued from the joint action of the other imposed constraints.

Recall that this version of the Tool requires that the precision constraints take the form of confidence interval half-widths. Once the sample allocation solutions have been obtained, the confidence interval half-widths can be computed using the allocated number of observations. Because the actual variance constraints imposed are inequality constraints, the half-widths given the allocation solutions are less than or equal to those originally specified.

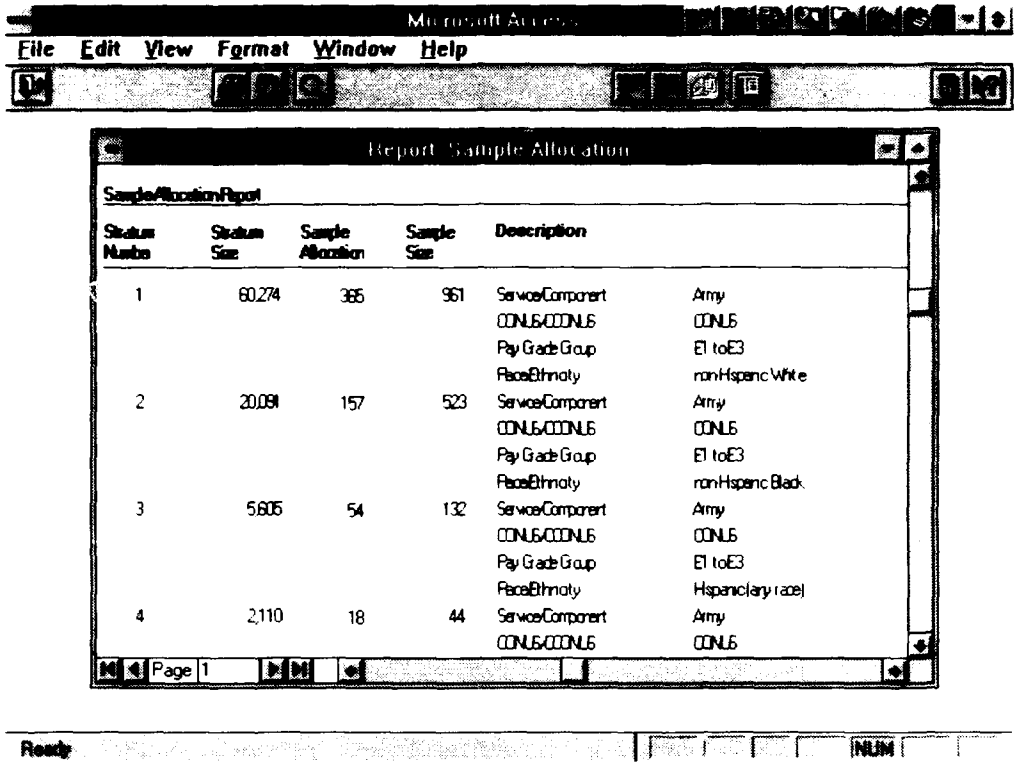
Often they are less than the original specification. A comparison of the *EXPECTED PRECISION* values on this form with the *PRECISION CONSTRAINT* values from the *Domain Data Report* will identify those domain estimates expected to have a higher level of precision than that specified for the design.

Design effects are shown for each of the domain estimates. The design effect measures the efficiency of the sampling design relative to a simple random sample having the same number of observations. Values less than one indicate the design is more efficient than simple random sampling and values greater than one indicate a less efficient design. Efficiency in a design effect context compares the variances obtained for a given number of observations and excludes any consideration of cost.

Because even minor changes in the sampling domain specifications can alter the allocation solution, printing the report for possible comparison with future modifications is recommended.



Sample Allocation Report



The final report within the sampling tool is the *Sample Allocation Report*. The report lists the allocation solutions obtained by *STRATUM NUMBER*. Strata are identified in the *DESCRIPTION* column of the report which lists the dimensions of stratification and the levels within each dimension used in defining the stratum.

For each stratum the report shows the *STRATUM SIZE*, the allocation solutions obtained (*SAMPLE ALLOCATION*) and the *SAMPLE SIZE*. The *SAMPLE ALLOCATION* is the number of stratum-level observations needed to jointly satisfy the set of imposed precision constraints. The *SAMPLE SIZE* is the size of the stratum-level sample to be selected in order to obtain the required number of observations given the expected response rate information for the stratum. The total *SAMPLE ALLOCATION* and *SAMPLE SIZE* values are provided on the last line of the last page.

Again, because even minor changes in the study specifications can alter the allocation solution, printing the report for possible comparisons with future results following design modifications is recommended.

References

Chromy, J. R. (1987). Design optimization with multiple objectives. In *Proceedings of the Section on Survey Research Methods* (pp. 194-199). Alexandria, VA: American Statistical Association.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York: John Wiley & Sons.

Mason, R. E., Wheeless, S. C., George, B. J., Dever, J. A., Riemer, R. A. & Elig, T.W. (1995). Sample allocation for the status of the armed forces surveys. *In Proceedings of the Section on Survey Research, Volume I* (pp. 769 - 774). Alexandria, VA: American Statistical Association.

Mason, R. E., Kavee, J. A., Wheeless, S. C., George, B. J., Riemer, R. A., & Elig, T. W. (1996). *The 1995 Armed Forces Sexual Harassment Survey: Statistical methodology report* (Report No. 96-016). Arlington, VA: Defense Manpower Data Center. (DTIC/NTIS No. AD A323 943)

Kuhn, H. W., & Tucker, A. W. (1951). Nonlinear programming. In J. Neyman (Ed.) *Proceedings of the Second Berkley Symposium on Mathematical Statistics and Probability* (pp. 481-492). Berkley, CA: University of California Press.

Wheeless, S. C., Mason, R. E., Kavee, J. D., Elig, T. W., & Riemer, R. A. (1997). *Armed Forces 1996 Equal Opportunity Survey: Statistical methodology report* (Report No. 97-025). Arlington, VA: Defense Manpower Data Center.

Wolter, K. M. (1985). *Introduction to variance estimation*. New York: Springer-Verlag, Inc.

Appendix A: Table Definitions

This appendix provides the variable name and corresponding description of the key design parameters for the tables created by the sampling tool.

COST DATA:

Stratum Number	sequential identification number per stratum
Average Cost	cost model value
Coefficient 1	cost coefficient for cost model component #1
:	:
Coefficient *C* *	cost coefficient for cost model component #*C*; *C* = number of cost components "included" in cost model
Average Response Rate	average response rate (sum of response rates 1-*R*)
Response Rate 1	proportion of response rate obtained on mailing #1
:	:
Response Rate *R* *	proportion of response rate obtained on mailing #*R*; *R* = number of mailings

DOMAIN INFORMATION:

Domain Number	sequential identification number per sampling domain
Domain Size	relative domain size
Initial Lambda	initial value for Lagrange multiplier
Final Lambda	final value for Lagrange multiplier
Design Effect	design effect

DOMAIN KEY:

Domain Number	sequential identification number
Variable Number	maximum number of value codes per domain
Domain Variable 1	first variable used to create domain
Variable Value 1	first variable value code
:	:
Domain Variable *D*	*D*-th variable used to create domain; *D* = number of domains
Variable Value *D*	*D*-th variable value code; *D* = number of domains
Domain Label	domain label
Precision Constraint	precision constraint per domain
Prevalence	prevalence estimate per domain
Domain Size	domain size calculated from frame (source data)
New Domain	flag to indicate newly created or newly re-defined sampling domains

LOOK-UP TABLE:

Stratum Number	sequential identification number per stratum
Stratum Size	stratum size calcuated from frame (source data)
Sample Size	estimated sample size (Integer Solution * Average Response Rate)

LOOK-UP TABLE: (continued)

Variable 1	label for dimension #1
Level Code 1	level code for dimension #1
:	:
Variable *S*	label for dimension #*S*; *S* = number of dimensions
Level Code *S*	level code for dimension #*S*

MODEL ITEMS:

Number of Mailings	number of mailings sent to study subjects
Frame Construction	indicator for inclusion of frame construction component in cost
Sample Selection	model
Instrumentation	indicator for inclusion of sample selection component in cost model
Data Collection	indicator for inclusion of instrumentation component in cost model
Data Editing	indicator for inclusion of data collection component in cost model
Data Processing	indicator for inclusion of data editing component in cost model
Analysis	indicator for inclusion of data processing component in cost model
	indicator for inclusion of analysis component in cost model

NEW STRATUM KEY: (updated version of STRATUM KEY)

Dimension	sequential identification number (1-number of dimensions)
Level	level of stratification
Stratification Variable	name of stratification variable
Code	variable value code
Label	label for variable value code

NEW STRATUM SIZES: (collapsed version of STRATUM SIZES)

Stratum Number	sequential identification number per stratum
Stratum Size	stratum size calculated from frame (source data)
Dim 1	dimension #1 and level code identification
:	:
Dim *S*	dimension #*S* and level code identification; *S* = number of
Dim1Label	dimensions
:	label for dimension #1
Dim*S*Label	:
Collapse Stratum	label for dimension #*S*; *S* = number of dimensions
	number of stratum that current stratum has been collapsed into

SAMPLE ALLOCATION:

Stratum Number	sequential identification number per stratum
Stratum Size	stratum size calculated from frame (source data)
Allocation Solution	solution to allocation
Integer Solution	integer portion of Allocation Solution
Sample Size	estimated sample size (Integer Solution * Average Response Rate)

STRATUM KEY:

Dimension	sequential identification number (1-number of dimensions)
Level	level of stratification
Stratification Variable	name of stratification variable
Code	variable value code
Label	label for variable value code

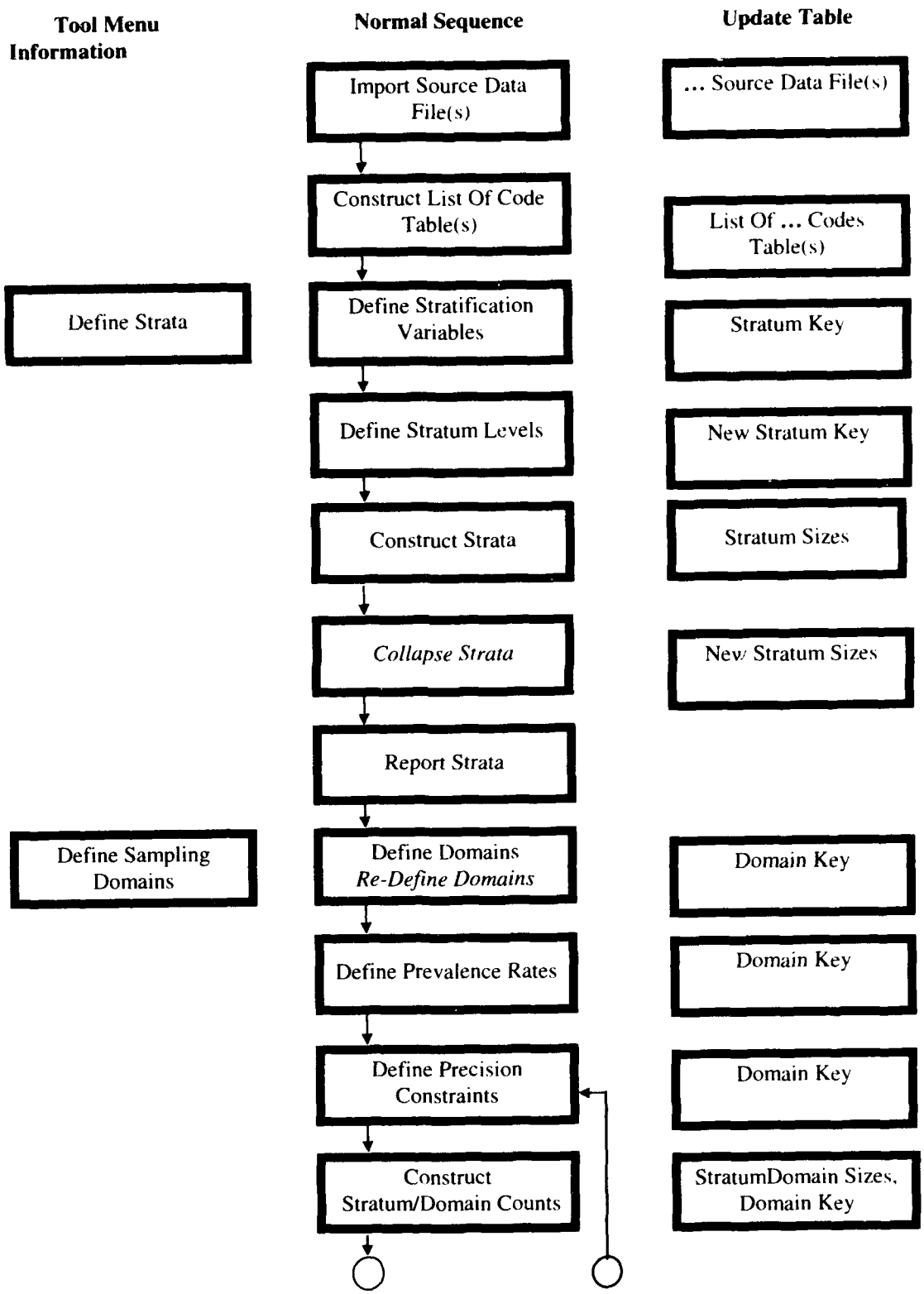
STRATUM SIZES:

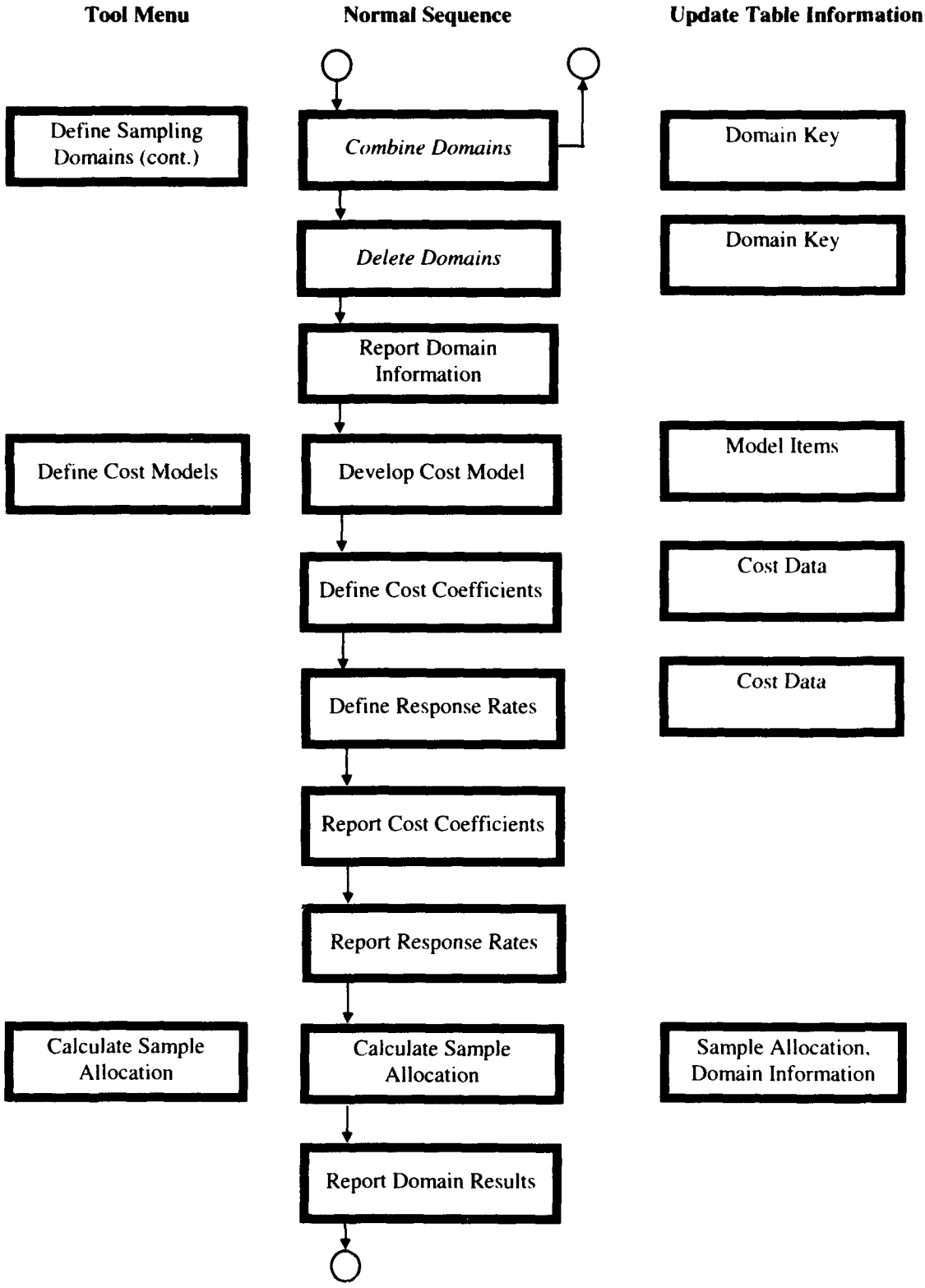
Stratum Number	sequential identification number per stratum
Stratum Size	stratum size calculated from frame (source data)
Dim 1	dimension #1 and level code identification
:	:
Dim*S* *	dimension #*S* and level code identification; *S* = number of
Dim1Label	dimensions
:	label for dimension #1
Dim*S*Label *	:
	label for dimension #*S*; *S* = number of dimensions

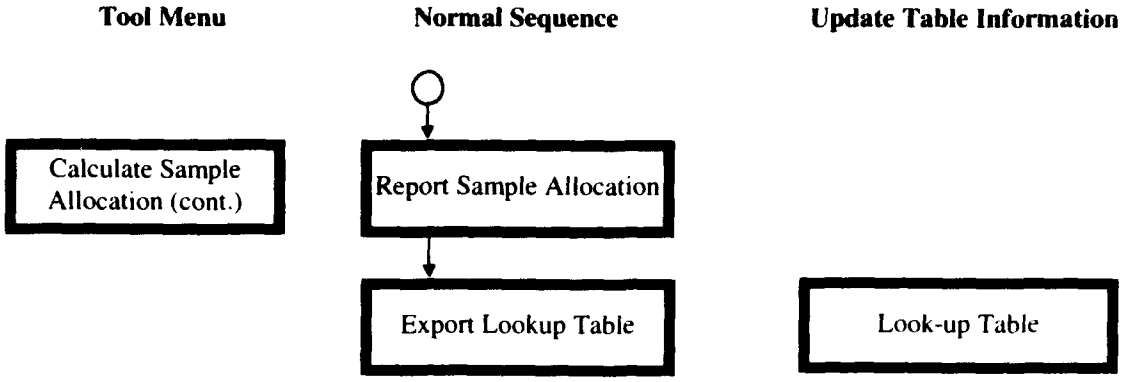
STRATUMDOMAIN SIZES:

Stratum Number	sequential identification number
Domain 1	size of domain #1 within stratum
Domain 2	size of domain #2 within stratum
:	:
Domain *D* *	size of domain #*D* within stratum; *D* = number of domains

Appendix B: Processing Steps




Tool Menu
Normal Sequence
Update Table Information
Define Sampling Domains (cont.)
Combine Domains
Domain Key
Delete Domains
Domain Key
Report Domain Information
Define Cost Models
Develop Cost Model
Model Items
Define Cost Coefficients
Cost Data
Define Response Rates
Cost Data
Report Cost Coefficients
Report Response Rates
Calculate Sample Allocation
Calculate Sample Allocation
Sample Allocation, Domain Information
Report Domain Results


Tool Menu
Normal Sequence
Update Table Information
Calculate Sample Allocation (cont.)
Report Sample Allocation
Export Lookup Table
Look-up Table

Appendix C: Technical Documentation

Mason, R. E., Wheeless, S. C., George, B. J., Dever, J. A., Riemer, R. A. & Elig, T.W. (1995). Sample allocation for the status of the armed forces surveys. *In Proceedings of the Section on Survey Research, Volume I* (pp. 769 - 774). Alexandria, VA: American Statistical Association.

SAMPLE ALLOCATION FOR THE STATUS OF THE ARMED FORCES SURVEYS

R. E. Mason, S. C. Wheeless, B. J. George, and J. A. Dever, Research Triangle Institute
R. A. Riemer and T. W. Elig, Defense Manpower Data Center
R. E. Mason, Research Triangle Institute, 3040 Cornwallis Road, Research Triangle Park, NC 27709

Key Words: Optimization, Sample Allocation

1. Introduction

The 1995 Status of the Armed Forces Surveys (SAFS) deal with gender, racial, and ethnic issues in the United States military establishment. A total of four surveys are involved. The examples used in this paper are drawn from one of the four surveys, known as the Form B survey, which deals with gender issues.

Each survey includes members of the four Armed Services, the Coast Guard, National Guard, and Reserves worldwide. Data collection is by mail. Sample individuals initially receive an introductory letter that explains the survey and solicits cooperation. The letter is followed by a package containing the questionnaire and instructions for completing and returning the information. The package is followed by a second letter thanking the individual for having returned the questionnaire or otherwise asking for its return. After a specified time has elapsed a second package containing the questionnaire and instructions is mailed to nonrespondents.

An unusual feature of these surveys is the large amount of information that is available for design purposes about the individuals that comprise the population. Not only are the demographics of individuals known in some detail, but so also are their occupational specialties, their work locations and settings, and their positions within the total organizational structure. This wealth of concomitant information is used to control the distribution of the sample for the purpose of providing predetermined levels of precision for estimates of parameters that describe key reporting domains.

The information is used to construct strata and to determine the sizes of the key reporting domains within each of the defined strata. Given the stratum sizes and their composition, variance constraints are placed on parameter estimates describing domains defined within one or more strata and overall. Equations are developed that describe the variances of the estimates and the variable survey costs in terms of the salient features of the design, which are constants in the equations, and the sample sizes to be allocated as specified by the design structure, which are the unknowns in the equations. The equations are solved simultaneously subject to the variance constraints to yield that allocation of the total sample that jointly satisfies the imposed variance constraints for the least cost.

This method for determining a sample allocation was first developed by J. R. Chromy for use in a medical provider record check survey conducted by the Research Triangle Institute in the late 1970s (Folsom et al. (1979)). The procedure is described in Chromy (1987).

The variance equations, of course, require knowledge of the relevant population variances. In practice the population variances are likely to be unknown, at least in advance of the survey, which is the case for these surveys. We have, as a consequence, defined the parameters of interest to be population proportions such that the (binomial) population variances are coincidentally specified with specifications for the values of the proportions. That is, the parameters of interest for determining the sample allocation are the relative sizes of specified key domains. The convention introduces some generality and provides a useful surrogate for other parameters. Certainly parameters describing other domain characteristics are unlikely to be reliably estimated if the domain sizes themselves cannot be. This choice of parameters is not restrictive if the requisite population variances are known

2. Sampling Design

A stratified random sampling design is used for the SAFS. Sample individuals are selected with equal conditional probabilities given the stratum and without replacement.

The dimensions of stratification are shown in Table 1 along with the maximum number of levels in each dimension. The dimension labeled as Unknown contains all individuals for which at least one of the variable values needed to identify the appropriate level of stratification is missing from the source files used to construct the sampling frame. The stratum sizes resulting from forming all possible crosses of levels within dimensions were computed and compared with the minimum stratum size consistent with a proportional allocation of a total sample size of 40,000.

Table 1. Dimensions And Levels Of Stratification

Dimension	Levels
Service	Army
	Navy
	Marine Corps
	Air Force
	Coast Guard
	Reserves and National Guard (AGR/TARS)
Location	Continental United States (CONUS)
	Outside Continental United States (OCONUS)
Pay Grade	Enlisted Grades E1-E4
	Enlisted Grades E5-E9
	Company Grade and Warrant Officers
	Field Grade Officers
Gender	male
	female
Race/Ethnicity	non-Hispanic White
	non-Hispanic Black
	Hispanic any race
	Other
Unknown	

Stratum cells smaller than the minimum were identified as candidates for collapsing into other cells

In undertaking the collapsing, the dimensions of stratification were considered to be nested in the order in which they are presented in Table 1. First, racial categories for females overall were collapsed into two levels, non-Hispanic White and Other, except for female Marine Corps officers stationed overseas for whom no racial categories were defined. Second, locations were collapsed within the Coast Guard and within the National Guard and Reserves combination. A total of 180 strata were constructed.

Key reporting domains at the level of the overall population were defined using the same variables and variable values as were used for stratification with one addition. The addition involved occupations, with domains defined by the representation of women in an occupation. Occupation specialties in the military are different for officers and enlisted personnel. In each case the relevant list of occupations was divided into quartiles based on the proportion of women. Within the first quartile, which might be described as the most extremely male dominated occupations, four domains were defined to further identify those occupations with the very lowest representation of women. Otherwise the domains were defined by the quartiles of the distribution, making a total of seven occupational domains.

The domain sizes used to allocate the sample are the gender specific proportions of persons reporting at least one of the behaviors that define unwanted sexual attention. Domains defined at the level of the overall population are termed main effect domains in what follows. First order interactions are defined by crossing pairs of main effect domains, for example, gender by race. Higher order interactions are similarly defined. In addition to being important in their own right, variance constraints imposed on main effect domains act to control unequal weighting effects induced by the total pattern of imposed constraints, particularly those imposed on the higher order interactions (i.e., smaller domains).

The number of main effect, first and second order interaction domains used to allocate the sample together with their associated variance constraints are shown in Table 2. The precision requirements cited in Table 2 are confidence interval half-widths.

3. Sample Allocation

The variance constraints take the form,

$$v_d(n_s) \le K_d \ , \qquad d = 1,2,\ldots,D$$

where $v_d(n_s)$ is the variance function for the d-th parameter estimate and K_d is the constraint imposed by the investigator. The form of the variance function is, of course, specified by the design. The notation is intended to suggest that, regardless of its form, the variance is a function of unknown sample sizes, n_s.

Table 2. Variance Constraints

Domain Description	Number of Domains	Precision Requirements
Gender	2	0.02
Location	2	0.03
Service [1]	6	0.05
Gender by Occupation	14	0.08
Gender by Race	8	0.05
Gender by Location	4	0.03
Gender by Service [1]	12	0.05
Females by Pay Grade Group [3]	6	0.03
Females, Enlisted by Service [1]	6	0.05
Females, Commissioned and Warrant Officers by Service [1]	6	0.05
Females, E1-E3, by Active Duty Service [2]	5	0.05
Females, E4, by Active Duty Service [2]	5	0.10
Females, E5-E6 by Active Duty Service [2]	5	0.10
Females, E7-E9 by Active Duty Service [2]	5	0.10
Females, Company Grade Officers by Active Duty Service [2]	5	0.10
Females, Field Grade Officers by Active Duty Service [2]	5	0.10
Males by Pay Grade Group [3]	6	0.05
Males, Enlisted by Service [1]	6	0.06
Males, Commissioned and Warrant Officers by Service [1]	6	0.06
Males, E1-E3 by Active Duty Service [2]	5	0.06
Males, E4-E9 by Active Duty Service [2]	5	0.06
Total	124	

[1] Army, Navy, Marine Corps, Air Force, Coast Guard, National Guard and Reserves

[2] Army, Navy, Marine Corps, Air Force, Coast Guard

[3] E1-E3, E4, E5-E6, E7-E9, Company Grade Officers, Field Grade Officers

In addition to the variance function, a cost function $c(n_s)$ is developed to describe the total variable cost of the survey in terms of the same unknown sample sizes. Variable costs may, in general, be both domain and stratum specific. The cost modeling exercise is, therefore, to develop equations that describe the domain and stratum costs as appropriate and then combine them in the proper proportions to obtain the overall cost.

Given the cost and variance functions, interest lies in determining the values ${}^*n_s$ that minimize the objective function,

$$o(n_s, \lambda_d) = c(n_s) + \sum_d \lambda_d (v_d(n_s) - K_d),$$

where the λ_d are generalized Lagrange multipliers, one for each of the variance constraints imposed. Taking derivatives of the objective function yields equations of the form,

$$-\frac{\partial o(c(n_s))}{\partial (n_s)} = \sum_d \lambda_d \frac{\partial o(v_d(n_s))}{\partial (n_s)}. \qquad [1]$$

If the variance constraints hold, then at ${}^*n_s$ there must exist values of the Lagrange multipliers ${}^*\lambda_d$ such that equation [1] evaluated at ${}^*n_s$ is true and additionally,

$$v_d({}^*n_s) \le K_d, \qquad [2]$$

$${}^*\lambda_d \ge 0, \qquad [3]$$

$${}^*\lambda_d (v_d({}^*n_s) - K_d) = 0. \qquad [4]$$

Equations [1] through [4], with ${}^*n_s$ substituted in equation [1], are the Karush-Kuhn-Tucker necessary conditions (Kuhn and Tucker (1951)). Sufficiency is

argued on the basis that the cost function $c(n_s)$ is a convex function and the constraints $K_d - v_d(n_s)$ are concave functions (see, for example, Hillier and Lieberman (1974), pages 722 through 725).

3.1 Variance Model

Define the indicator variables

$\delta_{d,h,i} = 1$, if the i-th individual in the h-th stratum belongs to the d-th domain,
$= 0$, otherwise,

$\delta_{h,i} = 1$, if the i-th individual in the h-th stratum reports having experienced at least one of the behaviors defining unwanted sexual attention,
$= 0$, otherwise.

Then the total members of the domain who report having experienced at least one of the behaviors is the quantity

$$N_d P_d = \sum_h \sum_{i=1}^{N_h} \delta_{d,h,i} \delta_{h,i}$$

where $i = 1, 2, \ldots, N_h$ identifies the individuals classified into the h-th stratum. The relative domain size is the population proportion

$$P_d = \sum_h \frac{N_h}{N_d} P_{d,h}$$

where

$$P_{d,h} = \frac{1}{N_h} \sum_{i=1}^{N_h} \delta_{h,i} \delta_{d,h,i} .$$

Denote the sample estimate of the proportion by

$$\hat{P}_d = \sum_h \frac{N_h}{N_d} \hat{P}_{d,h} .$$

with variance

$$v(n_s) = Var\{\hat{P}_d\} = \sum_h \left(\frac{N_h}{N_d}\right)^2 Var\{\hat{P}_{d,h}\}$$

where, if the stratum-level samples are selected with equal probability and without replacement,

$$Var\{\hat{P}_{d,h}\} = \left(\frac{N_h - n_h}{N_h - 1}\right)\left(\frac{P_{d,h}(1 - P_{d,h})}{n_h}\right) . \qquad [5]$$

At the level of an individual domain, the variance constraints in Table 2 are of the form

$$Var\{\hat{P}_d\} \le K_d = \left(\frac{CI\{\hat{P}_d\}}{1.96}\right)^2$$

where $CI\{\hat{P}_d\}$ are the confidence interval half-widths reported in Table 2.

3.2 Cost Model

A candidate list of activities to be potentially included in a cost model consists of the following items:

- sampling frame construction
- sample selection
- instrument development
- data collection
- data editing
- data processing
- data analysis and reporting

For the SAFS, with a single stage of sampling, variable survey costs are largely if not quite completely defined by the data collection, data editing, and data processing activities. Cost coefficients can be developed for these activities in terms of the per unit cost of packages sent out on the first and second mailings, C_1 and C_2, and on the per unit costs of packages that are returned, C_3.

The cost model takes the form,

$$c(n_s) = \sum_h n_h \bar{C}_h$$

where, denoting the response rates to the first and second mailing by R_1 and R_2 respectively,

$$\bar{C}_h = \frac{C_{1,h} + (1 - R_{1,h}) C_{2,h} + (R_{1,h} + R_{2,h}) C_{3,h}}{R_{1,h} + R_{2,h}} .$$

The h-subscripts allow the cost coefficients and response rates to be different in different strata if appropriate. Military postal services are used for these surveys such that the cost coefficients are the same in

all strata. However response rates were allowed to be different according to Service, pay grade, gender, race and ethnicity based on current experience with related surveys.

3.3 Allocation Solutions

Taking derivatives of the objective function with respect to the stratum-level sample sizes, equating to zero, and solving for the values n_h yields solutions of the form,

$$n_h = \sqrt{\frac{\sum_d \lambda_d \left(\frac{N_h}{N_d}\right)^2 \left(\frac{N_h}{N_h - 1}\right) P_{d,h}(1 - P_{d,h})}{\bar{C}_h}}.$$

The solutions ${}^*n_h$ and ${}^*\lambda_d$ are found numerically. If to start the numerical procedure the initial values of the Lagrange multipliers are set to

$$\sqrt{{}^0\lambda_d} = \frac{\sum_h \left(\frac{N_h}{N_d}\right)\left(\sqrt{P_{d,h}(1 - P_{d,h})}\right)\left(\sqrt{\bar{C}_h}\right)}{K_d},$$

then a comparison of the initial values ${}^0\lambda_d$ and the final values ${}^*\lambda_d$ will identify those variance constraints that exert the major influence in determining the sample allocation and, by implication, the cost of the survey. In general the initial values chosen for this purpose are those values of the Lagrange multipliers that satisfy the constraints individually. Then final values that are closest to these initial values identify those constraints that are the most important in determining the allocation. A small relaxation of the identified constraints can yield important reductions in the variable cost of the survey should the initially imposed constraints prove unaffordable. Constraints that are satisfied coincidentally with the imposition of other constraints will have final Lagrange multiplier values of zero.

4. Results

The variance constraints listed in Table 2 were determined over several iterations. Our initial specifications of the constraints proved too restrictive to be practical. At each iteration, those constraints that were the major determinants of the allocation solutions were identified and progressively relaxed until a set of constraints were developed that provided both an informative and an affordable study. Given the specifications in Table 2, the ten constraints that were the major determinants of the final allocation solutions are listed in order in Table 3.

Note that all of the constraints in Table 3 are second order interactions. This result is not surprising in that such constraints involve quite fine subdivisions of the total population. The first order interaction that is the most important in determining the sample allocation is that imposed on female field grade officers, with a Lagrange multiplier ratio of 0.8820. By comparison, all of the main effect constraints have ratios that are essentially zero, indicating that the constraints were coincidentally satisfied with the imposition of the other constraints.

Because the imposed constraints are inequality constraints, the average performance of the sample in general tends to be better, that is, tends to have smaller variances, than is suggested by the constraints themselves. Table 4 reports the range of confidence interval half-widths computed using the allocation solutions for comparison with the requirements listed in Table 2.

Shown also in Table 4 are the ranges of the design effects for the domain estimates. The major component of the design effect is, of course, the unequal weighting effect associated with the disproportionate sample allocation.

Design effects judged to be excessively large provide some guidance for modifying either the design strata or the domain constraints or both. For example, the Service associated with the design effect of 7.51 in Table 4 is the Coast Guard. The efficiency of the design for this main effect constraint could perhaps be improved by removing racially defined strata, as was done for the female Marine Corps stationed overseas, and collapsing pay grade strata. Alternately, or in addition, the variance constraints imposed on the Coast Guard higher order interaction domains could be relaxed.

5. References

Chromy, J. R., 1987, Design Optimization With Multiple Objectives, Proceedings of the Section on Survey Research Methods, American Statistical Association, p. 194-199.

Table 3. List of Ten Most Restrictive Constraints

Domain Description	$\frac{\lambda_d}{\lambda_d}$
Females, Field Grade Officers, Coast Guard	0.9964
Females, E7-E9, Coast Guard	0.9955
Females, E1-E3, Coast Guard	0.9932
Males, E1-E3, Coast Guard	0.9883
Males, Officers, Coast Guard	0.9852
Females, Field Grade Officers, Marine Corps	0.9795
Males, E1-E3, Air Force	0.9561
Females, E1-E3, Marine Corps	0.9401
Males, Officers, AGR/TAR	0.9191
Males, Officers, Marine Corps	0.9126

Table 4. Variances and Design Effects

Domain Description	Precision	Design Effects
Gender	0.009 to 0.014	1.34 to 2.01
Location	0.014 to 0.027	3.82 to 5.54
Service	0.022 to 0.042	3.96 to 7.51
Gender by Occupation	0.021 to 0.080	1.66 to 4.07
Gender by Race	0.012 to 0.050	1.13 to 2.71
Gender by Location	0.014 to 0.027	1.10 to 2.11
Gender by Service	0.016 to 0.050	1.06 to 1.71
Females by Pay Grade Group	0.012 to 0.030	1.48 to 2.02
Females, Enlisted by Service	0.019 to 0.029	1.00 to 1.49
Females, Commissioned and Warrant Officers by Service	0.020 to 0.040	1.00 to 1.09
Females, E1-E3 by Active Duty Service	0.046 to 0.050	1.31 to 1.63
Females, E4 by Active Duty Service	0.048 to 0.077	1.63 to 1.92
Females, E5-E6 by Active Duty Service	0.023 to 0.032	1.17 to 1.23
Females, E7-E9 by Active Duty Service	0.050 to 0.086	1.78 to 1.88
Females, Company Grade Officers by Active Duty Service	0.027 to 0.037	1.30 to 1.44
Females, Field Grade Officers by Active Duty Service	0.046 to 0.087	1.67 to 1.78
Males by Pay Grade Group	0.029 to 0.050	1.50 to 1.80
Males, Enlisted by Service	0.029 to 0.060	1.01 to 1.11
Males, Commissioned and Warrant Officers by Service	0.053 to 0.059	1.00 to 1.01
Males, E1-E3 by Active Duty Service	0.059 to 0.060	1.12 to 1.24
Males, E4-E9 by Active Duty Service	0.036 to 0.060	1.11 to 1.90

Folsom, R. E. Jr., J. R. Chromy and R. L. Williams, 1979, Optimum Allocation of a Medical Care Provider Record Check Survey: An Application of Survey Cost Minimization Subject to Multiple Variance Constraints, presented at the Joint National Meetings of the Institute of Management Science and the Operations Research Society.

Hillier, F. S., and G. J. Lieberman, 1974, Operations Research, Second Edition, Holden-Day, Inc., San Francisco CA, 800p.

Kuhn, H. W., and A. W. Tucker, 1951, Nonlinear Programming, Proceedings of the Second Berkley Symposium on Mathematical Statistics and Probability, J. Neyman ed., University of California Press, Berkeley, CA, p. 481-492.

Appendix D: Report Documentation Page